# به نام خدا

# **سوال اول** ) پاسخ به پرسمان

با استفاده از رابطه tf-idf امتیاز کلمات را در هر مستند محاسبه می کنیم . هر چه قدر یک کلمه در یک مستند بیشتر تکرار شوند مقدار tf آن بیشتر است و هر چه قدر یک کلمه در تعداد مستندهای بیشتری آمده باشد مقدار idf آن کلمه کمتر است به عبارتی دیگر هر چه یک کلمه خاص تر باشد مقدار idf برای آن کلمه بیشتر است . پس کلمات خاص تری که در یک مستند بیشتر تکرار شوند ؛ در کل وزن بیشتری دارند و شانس بیشتری برای قرار گرفتن در رتبه های اول دارند .

به ازای تمام کلمات موجود در یک مستند ، امتیاز آن کلمه را در آن مستند محاسبه می کنیم . بدین ترتیب بردار امتیاز حاصل می شود ؛ از اندازه این بردار برای نرمال سازی مقدار امتیاز کل هر مستند با بردار پرسمان استفاده می کنیم. میزان شباهت با بردار پرسمان را از طریق به دست آوردن زاویه برداری محاسبه می کنیم .

# الف) پرسمان ساده و متداول تک کلمه ایی

## پرسمان : فوتبال

همانطور که مشاهده می شود در تمامی اسناد بازیابی شده کلمه فوتبال مشاهده می شود و موضوع این اسناد خبری در مورد فوتبال می باشد پس می توان نتیجه گرفت که با پرسمان کاربر مرتبط است.

• ) پرسمان چند کلمه ایی متداول

پرسمان: لیگ برتر کشتی

```
rank: 1 , doc_id: 6054 , score: 8.3151

title: نميراده از رياست ليگ کشتي استعال نميراده از رياست ليگ کشتي استعال نميراده از رياست ليگ کشتي الله: المتعال نميراده از رياست ليگ کشتي الله: المتعال نميراده از رياست ليگ کشتي گيران و مربيان شاهد اين امر است. فدراسيون کشتي با توجه به توانمندي هاي نميراده همچنان از او به عنوان برادر و مشاور در کيار خود بهره خواهد برد ...

rank: 2 , doc_id: 3532 , score: 8.3845

title: نميره موراه از ديم الله بيك به اختلاف با رئيس کميته ملي الميبک به اختلاف با رئيس فدراسيون کشتي فرنو کشتي داره و الميک ده اختلاف با رئيس فدراسيون کشتي حل شد، پاسخ داد ...

location : 1 , اختلاف بياض بود الميبک درباره اينکه آيا اختلاف با عليرها ديبر، رئيس فدراسيون کشتي حل شد، پاسخ داد ...

rank: 3 , doc_id: 4589 , score: 8.2944

title: موراه نميک مينه ملي کشتي فرنگي منمون شد ...

uRl: https://www.farsnews.ir/news/14891025808661/م-بمنوب-شد/کشين-فرنگي، ماحب عناورين مختلف بيد شده است ...

rank: 4 , doc_id: 3265 , score: 8.2943

title: موراه نميک بردر ده هاي ملي کشتي فرنگي، ماحب عناورين مختلف بيد شده است ...

uRl: https://www.farsnews.ir/news/1489111808235/ موراه داد خام تختي اعلام سند ...

content: [' , اجتلاف بر ملي کشتي فرنگي، ماحب عناوره يو مختلف بيد هاي ملي کشتي فرنگي، ماحب عناوره يو مختلف بيد شده است ...

content: [' , اجتلاف بر ملي کشتي براي حمور در رفايت هاي فهرما بي آسيا (مغولستان)، باري هاي آسيابي (چين) و رفايت هاي براي مصور در رفايت هاي فهرما بي آسيا (مغولستان)، باري هاي آسيابي (چين) براي مسابقات فاره اي اعلام شد ...

content: [' , اختلاف براي در صدر وار دار دار دار دار داره دسا بقات فاره اي اعلام شد ... ۱۹ دير اين مسابقات فاره اي اعلام شد ... ۱۹ دير اين در صدر ...

content: [' , در مدر وار دار دار دار دار دار داره دسا بقات فاره اي اعلام شد ... ۱۹ دير اين مسابقات فاره اي اعلام شد ... ۱۹ دير اين مسابقات فاره اي اعلام شد ... ۱۹ دير اين مسابقات فاره اي داره در دمد فرار دارد ... ۱۹ دير اعدود دي ساروي در صدر فرار دارد ... ۱۹ دير اعدود دي ساروي در صدر فرار دارد ... ۱۹ دير اعدود دي ساروي در صدر فرار دارد ... ۱۹ دير اعدود دي ساروي در صدر فرار دارد ... ۱۹ دير اعدود دي ساروي در مدر فرار دارد ... ۱۹ دير ۱۹ دي د
```

```
المراح عبد المودد المراحة الراحة عبد المودد المراحة المراحة
```

در تصویر بالا صحت اسناد بازیابی شده ارزیابی شده است . همچنین به وضوح مشخص است که تمامی اسناد با پرسمان مرتبط هستند. در تمامی اسناد کلمه کشتی دیده می شود . همچنین کلمه برتر و لیگ نیز در رتبه های اول ظاهر شده اند .

پ) پرسمان تک کلمه کم تکرار

پرسمان: زيمباوه

همانطور که مشاهده می شود کلمه زیمباوه فقط در یک مستند آمده است که صحت آن در تصاویر بالا قابل مشاهده است .

ت) پرسمان کم تکرار چند کلمه ایی

پرسمان : اروشلیم صهیونیست

```
rank: 1 , doc_id: 6330 , score: 0.3836
```

به گرارش خبرگراری فارس، سنادیکای مقابله با غادی ساری روابط با رژیم سهیوتیستی در امارات با انتظار سنانیه او ۱۳۵ ککوم کرد، کاروان شنان این رژیم در مسابقات جهانی به میزنایی امارات حضور پیدا کرده است، به همین حاطر این سنادیکا به حضور ورزشکاران سهیونیست واکنش نشان داده است. رژیم سه نی این روزها با کمک دولت امارات در تلاش است تا روابط سیاسی، اقتمادی، ورزشی و ... را با کشورهای دیگر غادی کند، امارات سردمدار غادی ساری روابط با رژیم کودک کش سهیونیست نی این روزها با کمک دولت امارات در تلاش است تا روابط سیاسی، اقتمادی، ورزشی و ... را با کشورهای دیگر تنام فواد در حال خیاند به آرمان مفاوست فلسطین است. انتهای

#### تمامی مستندات بازیابی شده فقط کلمه صهیونیست را دارند ولی کلمه اورشلیم را ندارند.

### پرسمان: اورشلیم زیمباوه

### همانطور که مشاهده می شود فقط در یک مستند کلمه زیمباوه وجود دارد . صحت آن در تصاویر بالا بررسی شده است .

يرسمان: نيوتون گاليله

```
query: 4/15
O documents found that have at least 0.0001 consie score with user query
The results are ranked in the following order
--- response time = 0.001 seconds ---
```

```
project|⇒ cat <u>IR_data_news_12k.json</u> | jq | grep 'گالیله' -B2
project|⇒ cat <u>IR_data_news_12k.json</u> | jq | grep 'نیوتون' -B2
project|⇒ [
```

همانطور که مشاهده می شود در هیچ مستندی این پرسمان وجود ندارد .

### یرسمان : کنگره ضدتروریست

```
rank: 1, doc_id: 10586, sonre: 0.4413

title: برای درباره مسائل مشترک اتجادیه اروپا و ایران در نفستی با رئیس کنگره اسپانیا 1081: https://www.farsness.ir/news/46000000000076/0/20007-1 و عجو جرب سوسپالیست کا تالونیاست و از سال ۲۰۱۹ رییس کنگره اسپانیا 1081: https://www.farsness.ir/news/4600000000007/0/20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007-20007
```

```
به گرارش جوره بارلماند تعرگراری فارس. احمد نادری رئیس شورای اجرای نخادیه بین المجالس جها ندر درنان هاید شدن کرد ازی فارس. احمد نادری رئیس شورای اجرایی انجادیه بین المجالس جها ندر مطلبت در ضععه همین توسید توسید توسید با درنس قدر است المجالس جها ندر و استدار فرد موسید توسید توسید توسید توسید توسید توسید توسید از این در نسوسها استدار از این از استدار توسید توسید توسید توسید توسید از این از استدار توسید توسی
```

همانطور که مشاهده می شود در همه اسناد بالا کلمه کنگره موجود است و موضوع غالب این کنگره ها ضدتروریست است در نتیجه با پرسمان کاربر مرتبط است.

#### سوال دوم )

پرسمان های ب و ت را با شاخص مکانی بررسی می کنیم .

پرسمان: لیگ برتر کشتی

### (المراقع الله المراقع الله المراقع المراق

همانطور که مشاهده می شود فقط دو مستند دارای عبارت لیگ برتر کشتی هستند و این دو مستند در نتیجه این پرسمان ظاهر شده اند . از طرفی موضوع اصلی مستند ۵۴۳۰ در مورد کشتی نیست ؛ این مستند با موضوع حمایت ایرانسل از ورزش هندبال بیان شده است ولی چون دارای عبارت لیگ برتر کشتی است در نتیجه ظاهر شده است . این در حالی که در روش شباهت کسینوسی این مستند در نتایج بالا ظاهر نشده است و مستنداتی که بازیابی شده اند بیشتر به پرسمان شباهت دارند. پس در این پرسمان موتورجستجوی برداری نتیجه بهتری هم از لحاظ صحت هم از لحاظ پوشش نسبت به روش مکانی دارد.

از طرفی در حالتی پرسمان چند کلمه ایی وارد شود موتورجستجوی فاز اول شود ؛ مستنداتی که دارای بیشترین تعداد کلمه های پرسمان را در نتایج اولیه ظاهر می شوند . تعداد مستندات بازیابی شده با تعداد اسناد بازیابی شده با موتورجستجوی فاز اول یکسان است ولی رتبه بندی های تفاوت دارند . به وضوح می توان مشاهده کرد که در روش برداری اسناد مرتبط تر در رتبه های کمتری قرار می گیرند و دقت بهتری دارد.

پس در هر دو حالت جستجوی عبارت و جستجوی کلمات ، نتیجه فاز دوم با پرسمان کاربر مرتبط تر است .

### پرسمان : کنگره ضدتروریست

```
query : "كنگره مدتروريست"

phrase result = {6929: [[344, 345]]}

1 documents found that have at least 1 word of user query

The results are ranked in the following order
{1: {6929: 1}}

rank: 1 , doc_id: 6929 , frequency: 1

title: توضيحات يك منبع آگاه درباره وقفه مذاكرات وين URL: https://www.farsnews.ir/news/14001222000450/
```

" ( التوسيحات بك التواقي الله التوسيح الكران ولي من الورهائي أحر ود نزديك شده است و براساس اغلام برحي منابع تا رسيد، به بنظر ميرسد مداكرات وبي مه روزهائي آخر خود نزديك شده است و براساس اغلام ميرضي منابع تا رسيد، به بنوافي رات بنادي بنادي تعايده و تنهيل المعتملية على المالية التوسيط المعتملية التوسيط المعتملية التعليد التعليد المعتملية المعتملية التعليد المعتملية التعليد المعتملية المعتملية التعليد التعليد المعتملية التعليد التعلي

همانطور که مشاهده می شود فقط یک مستند داری عبارت پرسمان است که در نتیجه روش مکانی ظاهر شده است . این در حالی است که خیلی از اسناد با این پرسمان مرتبط هستند ولی در نتیجه ظاهر نشده اند ولی در نتیجه روش برداری اسناد بیشتری در نتایج ظاهر شده اند که با پرسمان مرتبط هم هستند در نتیجه پوشش بهتری دارند . در این مورد نیز در روش برداری نتیجه بهتری حاصل شد.