

به نام خدا

گزارش فاز دوم پروژه

۱- الف) یک پرسمان از کلمات ساده و متداول تک کلمه ای:

```
query = 'فوتبال'  
query_tokens = preprocess([query], True, True)[0]  
scores = cosine_similarity(query_tokens, positional_index, len(contents), docs_norms)  
print_result(scores, k=5)
```

🔗 1466 title: نگو نام: نفتی ها بهترین بازی خود را انجام دادند/بازیکنان جدیدمان کیفیت بالای خود را نشان دادند
url: <https://www.farsnews.ir/news/14001204001165/>

81 title: ماجدی: فوتبال کشور به تغییرات نیاز دارد
url: <https://www.farsnews.ir/news/14001223000539/>

6690 title: امیدواری ملی پوش سابق فوتبال ساحلی بابت تغییرات در کادرفنی تیم ملی
url: <https://www.farsnews.ir/news/14000927000275/>

139 title: تقدیر مربی تیم فوتبال خلیج فارس از ماجدی بابت رسیدگی به شائبه تبانی در لیگ جوانان
url: <https://www.farsnews.ir/news/14001222000297/>

860 title: دیدار مدیر تیم ملی ایران با سفیر کره جنوبی/ هیان: امیدوارم مقابل ایران مساوی کنیم
url: <https://www.farsnews.ir/news/14001213000138/>

همانطور که از تیتراخ خبر مشخص هست همگی مربوط به فوتبال می باشند. به عنوان مثال در تیتراخ خبر اول کلمه فوتبال مشاهده نمی شود اما اگر خود خبر را بخوانیم بارها کلمه فوتبال را می بینیم:

به گزارش خبرگزاری فارس از خوزستان، جواد نکونام پس از پیروزی تیمش مقابل نفت مسجد سلیمان در نشست خبری حضور پیدا کرد و اظهار داشت: شناخت کاملی از حریف داشتیم و بازی های آنها را آنالیز کرده بودیم. امروز بهترین بازی خود را ارائه کردند. شکل فوتبال آنها خوب بود و مقاومت خوبی مقابل ما داشتند. اگر دقت می کردیم زودتر به گل می رسیدیم، آنها اصلاً ضدفوتبال بازی نکردند و به آنها خسته نباشید می گویم، فکر نمی کردم بازی اینقدر سخت باشد. نفت تیم قابل احترامی است، اما ما هم در زدن ضربات آخر بلدشانس بودیم.

وی ادامه داد: نیمه اول یک گل زدیم و در نیمه دوم هم گل دوم را زدیم و موقعیت های بیشتری برای زدن گل داشتیم. خدا را شکر که بردیم و این برد را به مردم تبریک می گویم. کیفیت بازیکنان هر روز بهتر می شود و باید خودمان را برای مسابقات پیش رو آماده کنیم.

نکونام درمورد بازیکنان جدید و عملکرد آنها خاطرنشان کرد: وقتی دو گل از حریف پیش می افتیم این موضوع روی ذهنیت بازیکنان تاثیر دارد. نکته مهم این است که بازیکنان جدید ما با توجه به زحماتی که بازیکنان سال قبل کشیده اند که البته از آنها تشکر هم می کنم، عملکرد خوبی داشتند. از اعضای هیأت مدیره و مدیریت شرکت فولاد هم تشکر می کنم که در فصل نقل و انتقالات بازیکنان مدنظرمان را جذب کرده ایم. بازیکنان جدید کیفیت خود را نشان دادند و این موضوع را قبلاً هم گفته بودم که آنها از بهترین ها هستند، اما حالا همه باید کمک کنند تا تیم موفق باشد.

سرمرنی تیم فوتبال فولاد خوزستان درخصوص تغییراتی که در فدراسیون فوتبال ایجاد شده است، تصریح کرد: به عنوان کسی که در فوتبال هستم فقط می توانم نظرم را بگویم. کاری به تغییرات و اینکه عزیزی خادم چه کاری کرده و چه کاری نکرده ندارم، اما امیدوارم فوتبال ما رو به جلو حرکت کند، چون اگر اینگونه شود برای همه خوب است.

نکونام افزود: تجربه نشان داده کسانی که مدیریت فوتبالی داشته اند و توانسته اند کاری بکنند در بدنه فوتبال بوده اند. ماجدی آمده و همه ما خوشحال هستیم، چون او هم مثل دادکان فوتبالی است. امیدوارم فوتبال ما رو به جلو حرکت کند. ماجدی را سال های زیادی است که می شناسیم؛ او رشد کرده و اگر قرار است در فدراسیون بماند که دوست داریم بماند امیدوارم تمام تصمیمات فوتبالی باشد. فدراسیون همه کاره فوتبال است و دوست داریم که میرشاد ماجدی عزیز موفق باشد، چون از بدنه فوتبال است و درد فوتبال را می داند.

ب) یک پرسمان از عبارات ساده و متداول چندکلمه ای:

```
query = 'تیم ملی فوتبال'  
query_tokens = preprocess([query], True, True)[0]  
scores = cosine_similarity(query_tokens, positional_index, len(contents), docs_norms)  
print_result(scores, k=5)
```

اسکوچیچ: مردم متوجه شده‌اند که می‌توانند هدایت تیم ملی را به من واگذار کنند/هاشمیان سواد اروپایی از فوتبال دارد
url: <https://www.farsnews.ir/news/14001222000329/>

مصاحبه فارس با کارشناس فوتبال آسیا | از میراث بزرگ کی‌روش و قدرت ایران با اسکوچیچ تا انقلاب برانکو در عمان
url: <https://www.farsnews.ir/news/14001124000522/>

نکونام: نفتی‌ها بهترین بازی خود را انجام دادند/بازیکنان چندبیمان کیفیت بالای خود را نشان دادند
url: <https://www.farsnews.ir/news/14001204001165/>

ماجدی: فوتبال کشور به تغییرات نیاز دارد
url: <https://www.farsnews.ir/news/14001223000539/>

امیدواری: ملی‌پوش سابق فوتبال ساحلی بابت تغییرات در کادرفنی تیم ملی
url: <https://www.farsnews.ir/news/14000927000275/>

همانطور که از تیتراژ اخبار مشخص است همگی درباره ی تیم ملی فوتبال است. بخشی از متن خبر اول را مشاهده می‌کنیم:

به گزارش خبرگزاری فارس، سرمربی تیم ملی فوتبال کشورمان در مصاحبه با نشریه ناسیونال کرواسی در خصوص مهمان نوازی ایران، علاقه مردم به فوتبال، عملکرد تیم ملی و تصویر واقعی از کشورمان صحبت کرد.

اسکوچیچ گفت: ایران یک کشور فوق‌العاده است و مردمانی به شدت مهمان‌نواز دارد. مردم عاشق فوتبال هستند و استعداد در این کشور در فوتبال موج می‌زند. به نظرم کار کردن با بازیکنان جوان باعث می‌شود که اتفاقات بزرگ در ایران رخ دهد. در آسیا، ما بزرگترین از این نمی‌توانیم بشویم زیرا بهترین تیم آسیا هستیم. شما این را می‌توانید از آمار، نتایج، فکت‌ها و حقیقت‌های موجود متوجه شوید. ایران به صورت متوالی راهی جام جهانی شده که گواه این موضوع است.

وی ادامه داد: پتانسیل در این کشور بسیار بالاست. استعدادهای زیادی در این کشور وجود دارد که اگر مربیان با بازیکنان جوان کار کنند، این پتانسیل به ظهور خواهد رسید.

سرمربی تیم ملی فوتبال کشورمان اظهار داشت: منطقی است که فوتبال اروپا بیشتر دیده شود ولی فوتبال در آسیا روز به روز در حال پیشرفت است. پس از سرمربیگری در تیم ریکا و اینتر بلک اسلوونی، من برای اولین بار به آسیا آمدم و به کویت رفتم. در باشگاه العربی کار کردم و سپس راهی عربستان سعودی شدم و یک باشگاه بزرگتر به نام النصر را در دست گرفتم. این تیم بزرگترین باشگاه در عربستان بود. پس از کار در باشگاه‌های ایرانی به سرمربی تیم ملی فوتبال این کشور انتخاب شدم.

اسکوچیچ گفت: تیم ملی ایران به شدت اهمیت بالایی دارد، نه فقط در آسیا بلکه در کل جهان. زیرا این تیم در رنکینگ بیست و یکم جهان جای گرفته است. برای فهمیدن این موضوع، بدانید که صربستان در رده بیست و پنجم دنیا جای دارد و روسیه در جایگاه سی و پنجم است. دیگر نیازی نیست که جایگاه تیم‌های ملی دیگر را بگویم. به نظرم مردم بازی‌های آسیا را نگاه نمی‌کنند چون این مسابقات دست کم گرفته می‌شود. ممکن است چنین مسابقاتی با این سطح در کشور کرواسی نیز برگزار شود.

وی اظهار داشت: در کشور ما (کرواسی)، این موضوع دست کم گرفته شده است. در کل مربیان بزرگی در آسیا

پ) یک پرسمان دشوار و کم تکرار تک کلمه ای:



```
query = 'واترپلو'  
query_tokens = preprocess([query], True, True)[0]  
scores = cosine_similarity(query_tokens, positional_index, len(contents), docs_norms)  
print_result(scores, k=5)
```



```
1388 title: تجلیل از خانواده شهید حسن نوفلاح با تقدیم مدال قهرمانی  
url: https://www.farsnews.ir/news/14001205000938/ -تجلیل از -خانواده -شهید -حسن -نوفلاح -با -تقدیم -مدال -قهرمانی  
-----  
5690 title: سرپرست فدراسیون شنا، شیرجه و واترپلو منصوب شد  
url: https://www.farsnews.ir/news/14001011000202/ -سرپرست -فدراسیون -شنا -شیرجه -و -واترپلو -منصوب -شد  
-----  
5022 title: رضوانی رئیس فدراسیون شنا ماند/ 3 رای سفید برای تنها کاندیدا  
url: https://www.farsnews.ir/news/14001020000253/ -رضوانی -رئیس -فدراسیون -شنا -ماند -3 -رای -سفید -برای -تنها -کاندیدا  
-----  
5013 title: برنامه‌های رئیس فدراسیون شنا برای 4 سال آینده/ رضوانی: تلاش می‌کنیم به اهدافمان در بازی‌های آسیایی برسیم  
url: https://www.farsnews.ir/news/14001020000394/ -برنامه‌های -رئیس -فدراسیون -شنا -برای -4 -سال -آینده -رضوانی -تلاش -می‌کنیم -به  
-----  
1300 title: گرفتن ۷۰ نمونه تست دوپینگ در ۸ رشته طی یک هفته  
url: https://www.farsnews.ir/news/14001207000351/ -گرفتن -۷۰ -نمونه -تست -دوپینگ -در -۸ -رشته -طی -یک -هفته  
-----
```

این کلمه کم تکرار است. همان طور که مشاهده می شود از تیتراخبار مشخص است که مربوط به واترپلو می باشند. بخشی از متن خبر اول را مشاهده می کنیم:

به گزارش خبرگزاری فارس، سید سعید میرمهدی کاپیتان و امیرهوشنگ همراهی سرپرست تیم واترپلو شهید نوفلاح عصر امروز(پنجشنبه) با حضور در منزل شهید حسن نوفلاح کاپ قهرمانی و مدال طلای این تیم در لیگ برتر را به خانواده این شهید والامقام تقدیم کردند.

در این مراسم کاوه احمدی مدیرکل روابط عمومی وزارت ورزش و جوانان به نمایندگی از دکتر حمید سجادی وزیر ورزش و جوانان، سید علیرضا محسنی معاون آکادمی ملی المپیک و امیرداود کمالی نماینده فدراسیون شنا حضور داشتند و از خانواده این شهید تجلیل به عمل آوردند.

ت) یک پرسمان دشوار و کم تکرار چندکلمه ای:

```
query = 'واکسن کرونا ایرانی'
query_tokens = preprocess([query], True, True)[0]
scores = cosine_similarity(query_tokens, positional_index, len(contents), docs_norms)
print_result(scores, k=5)
```

7937 title: مرندی: رهبر انقلاب دُز سوم واکسن کرونا را دریافت کرده‌اند
url: <https://www.farsnews.ir/news/14001117000930/>

11966 title: الهیان: وزارت بهداشت درباره علت عدم پیش خرید واکسن فخرآ توضیح دهد
url: <https://www.farsnews.ir/news/14000729000462/>

7228 title: رئیس: مسافرت با رعایت اصول بهداشتی بلااشکال است
url: <https://www.farsnews.ir/news/14001214000475/>

9835 title: نایب رئیس مجلس: دولت در تزریق واکسن و کاهش نگرانی‌های مردم شاهکار کرده است
url: <https://www.farsnews.ir/news/14000923000300/>

9736 title: تأکید مخیر بر حمایت از تولیدکنندگان داخلی واکسن-کرونا/
url: <https://www.farsnews.ir/news/14000924000889/>

این عبارت کم تکرار است. همانطور که میبینیم نتایج به دست آمده تا حد زیادی مرتبط است. مثلاً خبر اول که بالاترین امتیاز را دارد درباره این است که رهبر واکسن ایرانی زده است و خبر دوم درباره ی واکسن فخرآ است که یک واکسن ایرانی است. بخشی از متن خبر اول را می بینیم:

به گزارش حوزه سیاسی خبرگزاری فارس، علیرضا مرندی رئیس فرهنگستان علوم پزشکی و وزیر بهداشت دولت ششم در «شبکه سلامت» رسانه ملی ضمن دعوت از مردم برای دریافت دزهای واکسن کرونا، از دریافت دُز سوم واکسن کرونا توسط رهبر انقلاب خبر داد.

مرندی در این باره گفت: همانطور که مقام معظم رهبری سه نوبت واکسن‌شان را زدند، واکسن ایرانی هم زدند، برکت هم زدند، ما هم باید همین‌کار را بکنیم، آنهايي که واکسن نزدند عجله کنند.

۲- نتایج قسمت ب و ت در فاز یک

(ب)

```
▶ query = 'تیم ملی فوتبال'  
res = search_query(query)  
print_output(res)
```

```
➡ Rank 1:  
title: ماجدی در نشست با صالحی امیری: امیدوارم در این دوره تیم المپیک موفق شود  
url: https://www.farsnews.ir/news/14001224000964/ ماجدی-در-نشست-با-صالحی-امیری-امیدوارم-در-این-دوره-تیم-المپیک-موفق  
-----  
Rank 2:  
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  
-----  
Rank 3:  
title: درخواست مدیر فوتبال کره جنوبی از مردم؛ بیاوید با هم ایران را شکست دهیم  
url: https://www.farsnews.ir/news/14001224000208/ درخواست-مدیر-فوتبال-کره-جنوبی-از-مردم-بیاوید-با-هم-ایران-را-شکست-دهیم  
-----  
Rank 4:  
title: ورزشگاه سنول پر از تماشاگر / دیدار ایران و کره جنوبی با 65 هزار تماشاگر  
url: https://www.farsnews.ir/news/14001224000189/ ورزشگاه-سنول-پر-از-تماشاگر-دیدار-ایران-و-کره-جنوبی-با-65-هزار-تماشاگر  
-----  
Rank 5:  
title: برگزاری نخستین جلسه کمیته المپیک با تیم امید فوتبال  
url: https://www.farsnews.ir/news/14001224000182/ برگزاری-نخستین-جلسه-کمیته-المپیک-با-تیم-امید-فوتبال  
-----
```

همان طور که مشاهده می‌کنیم در برخی موارد نتایج مرتبط گرفتیم ولی اخبار اول و دوم صرفاً به فوتبال و تیم مربوط بوده و درباره ی تیم ملی فوتبال نیست.

(ت)

```
query = 'واکسن کرونا ایرانی'
res = search_query(query)
print_output(res)
```

Rank 1:
title: شروط حضور هواداران در تمام ورزشگاه‌های کشور / مهدی: تغییر زمان دربی نهایی نشده است
url: <https://www.farsnews.ir/news/14001217000675/>

Rank 2:
title: نحوه توزیع سکوهای آزادی برای تماشاگران در دیدار ایران-عراق از زبان خبرنگار عراقی
url: <https://www.farsnews.ir/news/14001106000376/>

Rank 3:
title: وزیر کشور: ممکن است دوز سوم واکسن برای سفرهای خارجی اجباری شود
url: <https://www.farsnews.ir/news/14001217000520/>

Rank 4:
title: سخنگوی دولت: ۹۹ درصد سواحلی که در اختیار نهادهای دولتی بود آزاد شد
url: <https://www.farsnews.ir/news/14001212000835/>

Rank 5:
title: همایش علمی دانشگاهیان ایرانی خارج کشور | حمایت فارغ‌التحصیلان ایرانی دانشگاه‌های برتر دنیا برای بازگشت به کشور
url: <https://www.farsnews.ir/news/14001212000242/>

همان طور که مشاهده میکنیم نتایج تقریباً نامربوط اند.

نتایج بدست آمده برای پرسمان دشوار و کم تکرار، تقریباً نامرتب است. در فاز یک ما اسنادی را که حاوی کلمات پرسمان هستند را بازیابی میکنیم و اسناد را بر اساس اینکه در کدام سند تعداد بیشتری از کلمات پرسمان آمده است مرتب میکنیم. برای مثال در پرسمان «واکسن کرونا ایرانی» اسنادی را که کلمات واکسن»، «کرونا» و «ایرانی را دارند در رتبه های بالاتر قرار میگیرند و اسنادی که بعضی از این کلمات را ندارند در رتبه های بندی قرار میگیرند.

در فاز دوم ما تعداد تکرار یک کلمه در یک سند و اهمیت آن کلمه در کل اسناد را با استفاده از tf-idf پیدا و اسناد می کنیم و با استفاده از شباهت کسینوسی اسناد را از نظر شباهت مقایسه می کنیم. پس هر کلمه یک وزن دارد و اسناد بر اساس وزن و اهمیت کلمات موجود در پرسمان بازیابی و رتبه بندی میشوند. برای همین است که ما در فاز دوم نتایج دقیقتری گرفتیم و توانستیم به خوبی نتایج را رتبه دهی کنیم.