

# پروژه درس بازیابی اطلاعات

امیرحسین سرور — ۹۷۳۱۰۲۸

## گزارش فاز دوم

(۱) پاسخ به پرسمان در حالت‌های زیر. در هر مورد، تیتیر خبر بازیابی شده را به همراه جمله(هایی) که حاوی عبارت پرسمان بوده‌اند، گزارش کنید. همچنین در هر مورد با ذکر جزئیات شرح دهید که آیا سند بازیابی شده به پرسمان کاربر مرتبط هست یا خیر؟

الف) یک پرسمان از کلمات ساده و متداول تک‌کلمه‌ای

➤ پرسمان: انتخابات

```
>> Please Enter your Query: انتخابات
>> Top 10 Results for «انتخابات» :
=====
چکیده پروتکل های بهداشتی برگزاری انتخابات
پایان شمارش آراء هشتمین انتخابات سازمان نظام پزشکی در اکثر شهرها
قالیاف: حق الزحمه ناظران انتخابات شوراها این هفته پرداخت می‌شود
سه شنبه: برگزاری انتخابات هیات رئیسه مجلس در مجمع عمومی فراکسیون انقلاب اسلامی
گرم کردن رئیس فدراسیون بسکتبال برای 4 سال آینده!
ثبت نام ۲۹۶ داوطلب انتخابات میان دوره‌ای مجلس در حوزه انتخابیه تهران
بحث و بررسی پیرامون انتخابات هیات رئیسه مجلس و ریاست جمهوری در فراکسیون نیروهای انقلاب
ارائه گزارش وزیر کشور از فرآیند برگزاری انتخابات در کمیسیون امور داخلی کشور
پروتکل‌های ضدکرونا برای انتخابات فردا
انتخابات در ۲۴ شهر به صورت الکترونیکی برگزار می‌شود
=====
>> Retrieval Time: --- 5.340585231781006 seconds ---
=====
```



ب) یک پرسمان از عبارات ساده و متداول چندکلمه‌ای

➤ پرسمان: بانک رفاه کارگران

```
>> Please Enter your Query: بانک رفاه کارگران
>> Top 10 Results for «بانک رفاه کارگران» :
=====
بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت
رای کمیسیون اجتماعی به مجوز فعالیت وزارت رفاه در موضوعات بین المللی
حق بیمه اقشار ضعیف 2 برابر شده است/ وزیر رفاه ورود کند
بررسی برنامه وزیر پیشنهادی کار در کمیسیون آموزش
واکنش عضو کمیسیون اقتصادی مجلس به گرانی لوازم خانگی داخلی/ تولیدکنندگان بدانند کسی به آنها چک سفید امضاء نداده است
8 میلیون نفر بیمه درمانی ندارند/ انتقاد از وزارتخانه‌هایی که اطلاعات به یکدیگر نمی‌دهند
تصویب لایحه «مقاوله‌نامه حق تشکل و مذاکره دسته جمعی» در کمیسیون اجتماعی مجلس
تجمع تعدادی از کارگران پیمانکاری در پارس جنوبی
مدیریت «هفت تپه» به شرکت توسعه نیشکر و صنایع جانبی واگذار شد/ پرداخت معوقات و حقوق پرسنل و کارگران با همکاری بانکها
پیشنهاد ۴۵۰ هزار تومانی حق مسکن به دولت رفت
=====
>> Retrieval Time: --- 0.6582403182983398 seconds ---
=====
```

پ) پرسمان دشوار و کم تکرار تک کلمه‌ای

➤ پرسمان: ژیمناستیک

```
>> Please Enter your Query: ژیمناستیک
>> Top 10 Results for «ژیمناستیک» :
=====
دبیر مجمع فدراسیون ژیمناستیک مشخص شد
هشدار هیات ژیمناستیک تهران در خصوص سالن‌های مختلط و اقدامات غیراخلاقی
خیرخواه: برخی به دنبال فلج کردن ژیمناستیک هستند/ با بایکوت فدراسیون موفقیت‌ها بیشتر شد
تیت نام ۱۳ نامزد برای پست ریاست فدراسیون ژیمناستیک + اسامی
سوت‌زنی! تذکر و برخورد با سالن‌های ورزشی مختلط توسط وزارت ورزش
جزئیات تعطیلی‌های ورزش ایران تا ۹ مهر ۱۴۰۰/ کدام فعالیت‌های ورزشی در تهران ممنوع است؟
جزئیات تعطیلی ورزش ایران تا پایان تیرماه+ تصویر
دبیر: اگر من در مباحث فنی ۱۰ باشم، درستکار ۱۰۰ است/ بنا کاملاً بر اساس چرخه انتخابی عمل کرد!
=====
>> Retrieval Time: --- 0.7159090042114258 seconds ---
=====
```



ت) یک پرسمان دشوار و کم تکرار چندکلمه‌ای

➤ پرسمان: واکسن آسترازنکا

```
>> Please Enter your Query: واکسن آسترازنکا
>> Top 10 Results for «واکسن آسترازنکا» :
=====
محموله ۱.۴ میلیون دوزی واکسن کرونا وارد کشور شد
آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲۱ مرداد
آمار تفکیکی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه
اعلام آمار تفکیکی واکسیناسیون کرونا تا صبح امروز
اعلام آمار تجمیعی واکسیناسیون کرونا در ایران تا امروز
واکسن‌های کرونا با چه داروهایی تداخل دارند؟
اعلام آمار تفکیکی واکسیناسیون کرونا در کشور تا ۱۴ شهریور ماه
آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲ شهریور
تزریق بیش از ۳۱ میلیون دُز واکسن کرونا در کشور تا کنون
اعلام آمار تفکیکی واکسیناسیون کرونا تا ۸ شهریور ماه
=====
>> Retrieval Time: --- 0.6042900085449219 seconds ---
=====
```

برای هر کدام از پرسمان‌های بالا، جمله(هایی) را که حاوی پرسمان بوده‌اند برای اولین خبر بازیابی شده نشان می‌دهیم:

➤ پرسمان: انتخابات

➤ تیتیر خبر: چکیده پروتکل‌های بهداشتی برگزاری انتخابات

➤ متن خبر حاوی پرسمان:

چکیده پروتکل‌های بهداشتی برگزاری انتخابات ۱۴۰۰ اعلام شد.

به گزارش ایسنا، ستاد انتخابات کشور با همکاری وزارت بهداشت پروتکل‌های بهداشتی برای برگزاری انتخابات در شرایط وجود بیماری کرونا را تهیه کرده است تا سلامت رای دهندگان و مجریان انتخابات حفظ شود.

➤ ارتباط خبر به پرسمان: سند مربوط به مسائل بهداشتی در زمان برگزاری انتخابات است و ۴ بار کلمه انتخابات در آن تکرار شده است. سندی مرتبط با پرسمان ارزیابی می‌شود.



➤ **پرسمان:** بانک رفاه کارگران

➤ **تیتر خبر:** بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت  
**متن خبر حاوی پرسمان:**

در جلسه ستاد هماهنگی اقتصادی دولت، درباره ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران بحث و تصمیم‌گیری شد.

مجلس شورای اسلامی برای ساماندهی امور بانک رفاه کارگران، مقرر شد ضمن سرعت بخشیدن به بازپرداخت بدهی دولت به سازمان تامین اجتماعی و صندوق بازنشستگی کشور، اقدامات لازم برای بازپرداخت بدهی سازمان تامین اجتماعی به بانک رفاه کارگران نیز به عمل آید.  
 طرح "میزکار کارگزاری" بانک رفاه کارگران کلید خورد.

➤ **ارتباط خبر به پرسمان:** سند مربوط به بررسی وضعیت بانک رفاه کارگران است که با پرسمان کاربر هم مطابقت دارد و سندی مرتبط ارزیابی می‌شود.

➤ **پرسمان:** ژیمناستیک

➤ **تیتر خبر:** دبیر مجمع فدراسیون ژیمناستیک مشخص شد

➤ **متن خبر حاوی پرسمان:**

علیرضا باقری از سوی وزارت ورزش به عنوان دبیر مجمع ژیمناستیک انتخاب شد.  
 به گزارش ایسنا، ثبت نام از کاندیداهای پست ریاست فدراسیون ژیمناستیک از ۷ اردیبهشت ماه آغاز شده و به مدت ده روز کاری تا ۲۱ اردیبهشت ماه ادامه خواهد داشت.

➤ **ارتباط خبر به پرسمان:** سند مربوط به اتفاقی در مجمع فدراسیون ژیمناستیک است که سندی مرتبط به پرسمان محسوب می‌شود و در آن هم ۲ بار کلمه ژیمناستیک را مشاهده می‌کنیم.

➤ **پرسمان:** واکسن آسترازنکا

➤ **تیتر خبر:** محموله ۱,۴ میلیون دوزی واکسن کرونا وارد کشور شد

➤ **متن خبر حاوی پرسمان:**

معاون فنی گمرک ایران اعلام کرد که محموله‌ای حاوی ۱,۴ میلیون دوز واکسن کرونا امشب به ایران رسیده و تحویل وزارت بهداشت شده است.



مهرداد جمال ارونقی در گفت‌وگو با ایسنا تصریح کرد: محموله جدید واکسن کرونا ساعت ۱۰ از مبدا ایتالیا وارد فرودگاه امام خمینی شد که حاوی یک میلیون و ۴۰۰ هزار دوز واکسن آسترزنکا از مجموعه کوواکس است.

به گفته وی واکسن وارد شده همانند محموله‌های قبلی در حداقل زمان، تشریفات گمرکی را طی کرده و با حمل یکسره تحویل نمایندگان وزارت بهداشت شده است.

این مقام مسئول در گمرک ایران یادآور شد که بامداد یکشنبه نیز محموله یک میلیون دوزی دیگری از واکسن وارد و ترخیص شده است.

به گزارش ایسنا، پیش از دو محموله وارداتی امروز، از نیمه بهمن ماه بیش از ۳,۱ میلیون دوز واکسن از چهار کشور وارد شده بود.

این در حالی است که از اوایل اردیبهشت چندین محموله تجهیزات خط تولید واکسن نیز به کشور وارد شده و در اختیار ستاد اجرایی فرمان امام (ره) جهت تکمیل خط تولید واکسن قرار گرفته است.

➤ ارتباط خبر به پرسمان: سند مربوط ورود محموله واکسن آسترزنکا به کشور است که تطابق کاملی با کوثری کاربر دارد و سندی مرتبط ارزیابی می‌شود.

(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.

(ب) یک پرسمان از عبارات ساده و متداول چندکلمه‌ای

➤ پرسمان: بانک رفاه کارگران

```
>> Please Enter your Query: بانک رفاه کارگران
>> Top 10 Results for «بانک رفاه کارگران»:
=====
بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت
جزئیات گزارش تفریغ بودجه ۹۹ اکسری تراز عملیاتی بودجه ۶ برابر الزام برنامه ششم/۴۹ درصد احکام بودجه ۹۹ بر زمین ماند
گفت‌وگو با آر بی جی زن استقلال/از خاطرات با کلوب و حجازی تا انتقاد از مجیدی و گلی که باعث مرگ پدرش شد+فیلم
واکنش تند فدراسیون فوتبال به حواشی VAR و اجرای قرارداد با شرکت اسرائیلی
اطلاعیه باشگاه پرسپولیس در خصوص تبلیغات محیطی
بازگشت هیجان به والیبال/ جام سی‌وینجم به چه تیمی می‌رسد؟
اعلام آخرین اقدامات تراکتور برای بازشدن پنجره/ احتمال استفاده از بازیکنان جدید برابر گل‌گهر
والیبال باشگاه‌های آسیا| قهرمانی در غیاب شرقی‌های مدعی با تیمی پرستاره/ بزرگنمایی ممنوع!
واکنش مدیرعامل سابق استقلال به اختلاف ۳۰ میلیاردی و مزاد بودن غفوری
نبرد ایرانی‌ها در پلاس لیگا و درخشش سید در لیگ ترکیه/ از لژیونرهای والیبال چه خبر؟
=====
>> Retrieval Time: --- 0.7426419258117676 seconds ---
```



ت) یک پرسمان دشوار و کم تکرار چندکلمه‌ای

➤ پرسمان: واکسن آسترانکا

```
>> Please Enter your Query: واکسن آسترانکا
>> Top 10 Results for «واکسن آسترانکا» :
=====
امکان ایجاد لخته خون در واکسن آسترانکا چقدر است؟
محموله ۱.۴ میلیون دوزی واکسن کرونا وارد کشور شد
نکاتی که باید در مورد واکسناسیون کرونا بدانیم
محموله ۱.۴ میلیون دوزی واکسن کرونا وارد کشور شد
واکسن‌های کرونا با چه داروهایی تداخل دارند؟
واکسناسیون؛ عقلانی‌ترین راه مقابله با کرونا/پرهیز از تزریق واکسن خارج از چرخه عمومی واکسناسیون
نکاتی که باید در مورد واکسناسیون کرونا بدانیم
مهم‌ترین سلاح مبارزه با کرونا
مقررات تازه برای سفر زمینی ایران و ارمنستان
واکسناسیون؛ عقلانی‌ترین راه مقابله با کرونا/پرهیز از تزریق واکسن خارج از چرخه عمومی واکسناسیون
=====
>> Retrieval Time: --- 0.6666374206542969 seconds ---
```

➤ مقایسه نتایج روش مکانی و روش برداری:

همان‌طور که می‌بینیم، در قسمت ب اولین نتیجه‌ای که توسط هردو روش بازیابی شده، یکسان است و در آن صریحاً به «بانک رفاه کارگران» و مسائلی در آن رابطه پرداخته شده است و سندی مرتبط محسوب می‌شود. در قسمت ت نیز برای روش مکانی تیترا اولین سند بازیابی شده دقیقاً شامل کوئری کاربر است و در روش برداری نیز کوئری کاربر به طور دقیق در متن خبر وجود دارد. اما اگر کلی‌تر به نتایج نگاه کنیم، در روش مکانی نتایجی که در رنگ بالاتری قرار گرفته‌اند، عموماً اسنادی هستند که پرسمان کاربر را به طور کامل در خود دارند؛ یعنی مثلاً عبارت «بانک رفاه کارگران» یا «واکسن آسترانکا» را در متن یا تیترا خبر به صورت کامل می‌توان یافت. بعد از آن نتایجی آمده‌اند که تمرکز بیشتری بر روی کلمات کوئری به طور مستقل‌تری دارند؛ تا زمانی که تنها به وجود کلمات کوئری توجه می‌شود یعنی تنها کفایت که خبر بازیابی شده شامل «بانک»، «رفاه» یا «کارگران» و همچنین «واکسن» یا «آسترانکا» باشد تا بازیابی شود و در واقع اهمیتی به تاثیر و وزن هر کلمه داده نمی‌شود. البته در اینجا هم ۱۰ خبر بیشتر بازیابی نشده‌اند. در روش برداری، در ابتدا تمرکز بر روی اسنادی می‌رود که هم کلماتی از کوئری که وزن بیشتری دارند، مثل «آسترانکا» یا «رفاه»، در آن‌ها بیشتر به چشم بخورد (idf) هم اینکه تعداد بیشتری از کلمات کوئری را شامل باشند (tf). به همین علت مثلاً در اولین خبر بازیابی شده برای واکسن آسترانکا در روش برداری، ۸ بار کلمه واکسن به چشم می‌خورد در حالی که در خبر بازیابی شده در حالت مکانی، ۲ بار کلمه‌ی واکسن ظاهر می‌شود که در هر دو حالت هم در کنار کلمه‌ی آسترانکا قرار دارد.



- Mean Reciprocal Rank

برای مدل آموزش داده شده روی مجموعه اخبار داده شده:

تیتیر خبر	Reciprocal Rank score (1/K)	rank position of first relevant doc (K)	تعداد نتایج بازیابی شده	پرسمان	نوع پرسمان
جایجایی برنامه‌های رادیو و تلویزیونی نامزدها	0.5	2	10	انتخابات	ساده تک‌کلمه
بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت	1	1	10	بانک رفاه کارگران	ساده چندکلمه
هشدار هیات ژیمناستیک تهران در خصوص سالن‌های مختلط و اقدامات غیراخلاقی	0.16	6	10	ژیمناستیک	دشوار تک‌کلمه
آمار تولید، تحویل و تزریق واکسن کووایران برکت تا امروز	1	1	10	واکسن کووایران برکت	دشوار چندکلمه

$$\Rightarrow MRR = \frac{1}{4} (0.5 + 1 + 0.16 + 1) = 0.66$$

برای مدل آموزش داده شده از قبل:

تیتیر خبر	Reciprocal Rank score (1/K)	rank position of first relevant doc (K)	تعداد نتایج بازیابی شده	پرسمان	نوع پرسمان
سخنگوی جبهه اصلاحات: بارها گفته‌ایم در انتخابات ۱۴۰۰ کاندیدا نداریم	0.33	3	10	انتخابات	ساده تک‌کلمه
بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت	0.5	2	10	بانک رفاه کارگران	ساده چندکلمه
دبیر مجمع فدراسیون ژیمناستیک مشخص شد	0.25	4	10	ژیمناستیک	دشوار تک‌کلمه
اولویت تزریق واکسن آنفلوآنزا با چه کسانی است؟	1	1	10	واکسن آنفلوآنزا	دشوار چندکلمه

$$\Rightarrow MRR = \frac{1}{4} (0.33 + 0.5 + 0.25 + 1) = 0.52$$



• Precision@K

برای مدل آموزش داده شده روی مجموعه اخبار داده شده:

ان	پرسمان	تیتتر خبر	rank	relevant	precision@rank
ساده تک کلمه	انتخابات	کواکبیاں حتما در انتخابات ثبت نام می کند	1	Yes	1
		جابه جایی برنامه های رادیو و تلویزیونی نامزدها	2	Yes	1
		پایان شمارش آراء هشتمین انتخابات سازمان نظام پزشکی در اکثر شهرها	3	Yes	1
		اصلاح طلبان در انتخابات شوراها مشارکت پرشور دارند	4	Yes	1
		سخنگوی جبهه اصلاحات: بارها گفته ایم در انتخابات ۱۴۰۰ کاندیدا نداریم	5	Yes	1
ساده چند کلمه	بانک رفاه کارگران	بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت	1	Yes	1
		حق بیمه افشار ضعیف ۲ برابر شده است/ وزیر رفاه ورود کند	2	No	0.5
		مدیریت «هفت تبه» به شرکت توسعه نیشکر و صنایع جانبی واگذار شد/ پرداخت معوقات و حقوق پرسنل و کارگران با همکاری بانکها	3	No	0.33
		تامین اجتماعی و بازنشستگی کشوری در پرداخت حقوق بازنشستگان با مشکل مواجه اند	4	No	0.25
		بررسی برنامه وزیر پیشنهادی کار در کمیسیون آموزش	5	No	0.2
دشوار تک کلمه	ژیمناستیک	جزئیات تعطیلی های ورزش ایران تا ۹ مهر ۱۴۰۰ / کدام فعالیت های ورزشی در تهران ممنوع است؟	1	Yes	1
		جزئیات تعطیلی ورزش ایران تا پایان تیرماه	2	Yes	1
		صدای لالایی ایران خاموش شد	3	No	0.66
		درگذشت خالق نمایشگاه «بسم الله» به علت کرونا	4	No	0.5
		جشنواره تیرو کمان اصحاب رسانه قم برگزار شد	5	No	0.4
دشوار چند کلمه	واکسن آسترازنکا	آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲ شهریور	1	Yes	1
		تزریق بیش از ۳۱ میلیون دُز واکسن کرونا در کشور تا کنون	2	Yes	1
		اعلام آمار تفکیکی واکسیناسیون کرونا در کشور تا ۱۴ شهریور ماه	3	Yes	1
		اعلام آمار تفکیکی واکسیناسیون کرونا تا ۸ شهریور ماه	4	Yes	1
		آمار تولید، تحویل و تزریق واکسن کووایران برکت تا امروز	5	No	0.8

\* خانه هایی که پررنگ شده اند مقادیر  $Prec@1$  و  $Prec@5$  را نشان می دهند که در گزارش خواسته شده است.

➤  $Mean\ Average\ Precision = \frac{1}{4} (1 + 0.2 + 0.4 + 0.8) = 0.6$





- Precision@K

برای مدل آموزش داده شده از قبل:

ان	پرسمان	تیتر خبر	rank	relevant	precision@rank
ساده تک کلمه	انتخابات	کواکبیاں حتما در انتخابات ثبت نام می کند	1	Yes	1
		اصلاح طلبان در انتخابات شوراها مشارکت پرشور دارند	2	Yes	1
		سخنگوی جبهه اصلاحات: بارها گفته‌ایم در انتخابات ۱۴۰۰ کاندیدا نداریم	3	Yes	1
		آخرین خبر از حضور جهرمی در انتخابات ریاست جمهوری	4	Yes	1
		جابه جایی برنامه‌های رادیو و تلویزیونی نامزدها	5	Yes	1
ساده چند کلمه	بانک رفاه کارگران	مدیریت «هفت تپه» به شرکت توسعه نیشکر و صنایع جانبی واگذار شد/ پرداخت معوقات و حقوق پرسنل و کارگران با همکاری بانکها	1	No	0
		بررسی ساماندهی و بهبود وضعیت مالی بانک رفاه کارگران در جلسه ستاد هماهنگی اقتصادی دولت	2	Yes	0.5
		حق بیمه اقشار ضعیف ۲ برابر شده است/ وزیر رفاه ورود کند	3	No	0.33
		بیش از پنج میلیون بیکار داریم/ برای اشتغال‌زایی نیازمند کمک سایر وزارتخانه‌ها هستیم	4	No	0.25
		تامین اجتماعی و بازنشستگی کشوری در پرداخت حقوق بازنشستگان با مشکل مواجه اند	5	No	0.2
دشوار تک کلمه	ژیمناستیک	جزییات تعطیلی ورزش ایران تا پایان تیرماه	1	Yes	1
		جزییات تعطیلی‌های ورزش ایران تا ۹ مهر ۱۴۰۰/ کدام فعالیت‌های ورزشی در تهران ممنوع است؟	2	Yes	1
		۲۳ رشته ورزشی زیر تیغ پروتکل‌های کرونا	3	Yes	1
		دبیر مجمع فدراسیون ژیمناستیک مشخص شد	4	Yes	1
		خیرخواه: برخی به دنبال فلج کردن ژیمناستیک هستند/ بایکوت فدراسیون موفقیت‌ها بیشتر شد	5	Yes	1
دشوار چند کلمه	واکسن آسترازنکا	آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲ شهریور	1	Yes	1
		اعلام آمار تفکیکی واکسیناسیون کرونا تا ۸ شهریور ماه	2	Yes	1
		آمار تولید، تحویل و تزریق واکسن کووایران برکت تا امروز	3	No	0.66
		اعلام آمار تفکیکی واکسیناسیون کرونا در کشور تا ۱۴ شهریور ماه	4	Yes	0.75
		چه کسانی نباید دُز دوم واکسن کرونا را تزریق کنند؟	5	Yes	0.8

\* خانه‌هایی که پررنگ شده‌اند مقادیر  $Prec@1$  و  $Prec@5$  را نشان می‌دهند که در گزارش خواسته شده است.

➤  $\text{Mean Average Precision} = \frac{1}{4} (1 + 0.2 + 1 + 0.8) = 0.75$



➤ مقایسه عملکرد مدل در حالت بازنمایی tf-idf و بازنمایی word2vec :

الف) پرسمان از کلمات ساده و متداول تک کلمه‌ای

➤ پرسمان: انتخابات

```
>> Please Enter your Query: 🔍
>> Top 10 Results for «انتخابات» :
=====
کواکبیاں حتما در انتخابات ثبت نام می کند
جابه جایی برنامه های رادیو و تلویزیونی نامزدها + جدول
پایان شمارش آراء مشتمین انتخابات سازمان نظام پزشکی در اکثر شهرها
اصلاح طلبان در انتخابات شوراها مشارکت پرشور دارند
سخنگوی جبهه اصلاحات: بارها گفته ایم در انتخابات 1400 کاندیدا نداریم
نماینده تام الاختیار معنی در انتخابات مشخص شد
حضور رئیس و نماینده ولی فقیه در بنیاد شهید و امور ایثارگران پای صندوق رأی گیری
حضرتی رئیس ستاد انتخاباتی پزشکيان شد
تصميم اسحاق جهانگیری برای ثبت نام در انتخابات قطعی شد
دبیر مجمع فدراسیون ژیمناستیک مشخص شد
=====
>> Retrieval Time: --- 0.1047217845916748 seconds ---
=====
```

مدل آموزش داده شده روی مجموعه اخبار داده

```
>> Please Enter your Query: 🔍
>> Top 10 Results for «انتخابات» :
=====
کواکبیاں حتما در انتخابات ثبت نام می کند
اصلاح طلبان در انتخابات شوراها مشارکت پرشور دارند
سخنگوی جبهه اصلاحات: بارها گفته ایم در انتخابات 1400 کاندیدا نداریم
آخرین خبر از حضور جهری در انتخابات ریاست جمهوری
جابه جایی برنامه های رادیو و تلویزیونی نامزدها + جدول
کاندیداهای نباید تبلیغات زود هنگام را شروع کنند
جلسه بعدی جبهه اصلاح طلبان ایران یکشنبه برگزار می شود
روحانی و لاریجانی در مسیر اتحاد گفتگویی/ چرا اعتدال نمیخواهد جزئی از نظام باشد؟
امروز راهبرد انتخاباتی اصلاح طلبان برای انتخابات ریاست جمهوری بررسی می شود
آغاز بررسی صلاحیت داوطلبان انتخابات شوراهای شهر در هیات های نظارت شهرستان
=====
>> Retrieval Time: --- 0.10896849632263184 seconds ---
=====
```

مدل آموزش داده شده از قبل

همانطور که مشاهده می شود نتایج در حالت مدل آموزش داده شده از قبل و مدل آموزش داده شده بر روی اخبار تفاوت چندانی ندارند و هردو به نسبت عملکرد خوبی دارند.

نسبت به نتایج tf-idf در صفحه ۱، عملکرد تقریباً بهتری را شاهد هستیم؛ به این صورت که در نتایج tf-idf انتخابات به صورت عام در نظر گرفته شده و تمامی نتایج مربوط به هرا انتخاباتی بازبایی شده است، در حالی که در مدل تعبیه گذاری کلمه، تمرکز بر روی انتخابات عمده تری است که می تواند ناشی از توجه به کلمات اطراف یک کلمه نیز باشد.

ب) یک پرسمان از عبارات ساده و متداول چند کلمه‌ای

➤ پرسمان: وزارت نفت



```
>> Please Enter your Query: نفت نفت
>> Top 10 Results for «وزارت نفت» :
=====
تشکیل کمیته مشترک کمیسیون انرژی و وزارت نفت برای لایحه بودجه سال 1401 و برنامه هفتم توسعه
تشکیل کمیته مشترکی بین مجلس و وزارت نفت درباره لایحه بودجه سال ۱۴۰۱ و برنامه هفتم توسعه
حضور وزیر نفت در کمیسیون برنامه و بودجه
اساسنامه شرکت پالایش و پخش فرآورده‌های نفتی به هیئت دولت ارسال شده است
مأموریت کمیسیون بودجه به دیوان محاسبات برای ارائه اطلاعات هدفمندی بارانه‌ها
«نقض حاکمیت ملی عراق» با قراردادی 50 ساله میان اقلیم و ترکیه
کاهش 30 درصدی سطح مخازن نیروگاه‌ها نسبت به سال گذشته / تلاش وزارت نفت برای تأمین سوخت زمستانی
بررسی شکایت علیه وزارت نفت در کمیسیون اصل نود/تشکیل کمیته برای بررسی تخلفات
هفتمین نامه رئیس دیوان محاسبات/ اولویت‌های قانونی وزارت نفت به اطلاع وزیر رسید
ایران با سوریه تجارت دریایی می‌کند!
=====
>> Retrieval Time: --- 0.18411707878112793 seconds ---
=====
```

مدل آموزش داده شده از قبل

```
>> Please Enter your Query: نفت نفت
>> Top 10 Results for «وزارت نفت» :
=====
کاهش 30 درصدی سطح مخازن نیروگاه‌ها نسبت به سال گذشته / تلاش وزارت نفت برای تأمین سوخت زمستانی
تلاش وزارت نفت برای رفع نگرانی‌های تأمین گاز زمستان
وزیر نفت به رادیو اقتصاد می‌رود
تشکیل کمیته مشترک کمیسیون انرژی و وزارت نفت برای لایحه بودجه سال 1401 و برنامه هفتم توسعه
اختصاص ۲۰ میلیارد تومان به تیم نفت آبادان
تشکیل کمیته مشترکی بین مجلس و وزارت نفت درباره لایحه بودجه سال ۱۴۰۱ و برنامه هفتم توسعه
حضور وزیر نفت در کمیسیون برنامه و بودجه
بررسی ذخیره سازی سوخت زمستانی در کمیسیون انرژی/ مذاکره با ترکمنستان برای واردات گاز
نیروهای شرکتی در وزارت خانه‌ها تعیین تکلیف نشدند
موانع واگذاری سهام دولت در شرکت پالایش نفت امام خمینی (ره) شازند بررسی و رفع شد
=====
>> Retrieval Time: --- 0.10076522827148438 seconds ---
=====
```

مدل آموزش داده شده روی مجموعه اخبار داده شده

```
>> Please Enter your Query: نفت نفت
>> Top 10 Results for «وزارت نفت» :
=====
خداحافظی داریوش یزدی با بازیکنان نفت مسجدسلیمان
سرزمینی جدید نفت معارفه شد / فکری: به این تیم عرق دارم
پیوستن جاسم کرار به نفت مسجدسلیمان
اختصاص ۲۰ میلیارد تومان به تیم نفت آبادان
لژیونر خوزستانی به نفت مسجدسلیمان پیوست
سه پیشنهاد خارجی برای مسلم مجدی
جلسه مدیرعامل و سرزمینی نفت مسجدسلیمان با عزیزی‌خادم و بهاروند
پاسخ منفی پورموسوی به سایپا / ۱۲ میلیارد در یک قدمی باشگاه نفت
قرارداد منصوریان با نفت آبادان امضا شد / حضور او در تمرینات
نفت مسجدسلیمان با چک میلیونی رئیس شرکت وارد ورزشگاه شد
=====
>> Retrieval Time: --- 0.6106460094451904 seconds ---
=====
```

مدل tf-idf

در اینجا می‌توان دید که مدل tf-idf عملکرد بسیار ضعیفی دارد و تمرکز خود را بر روی کلمه‌ای که ظاهراً وزن بیشتری داشته است (نفت) گذاشته که این موضوع باعث بایاس شدن نتایج به سمت اخباری شده که دارای کلمه «نفت» هستند که حتی تیم‌های فوتبال را نیز شامل می‌شود. در نتایج word2vec، در هر دو مدل عملکرد نسبتاً خوبی را می‌بینیم که در مدلی که

با داده‌های بیشتری آموزش دیده، این عملکرد بهتر نیز می‌باشد (نتایج نامربوط کمتری وجود دارد)

پ) پرسمان دشوار و کم تکرار تک کلمه‌ای

➤ پرسمان: سینوفارم



```
>> Please Enter your Query: ایف
>> Top 10 Results for «سینوفارم» :
=====
واکنش بازیکن نفت مسجدسلیمان به صحبت‌های پیروانی: با شرافت بازی کردیم / این حرف‌ها زشت است
هفته دوم لیگ برتر پرسپولیس مقابل تیم سختکوش نساجی / آلومینیوم و فولاد به دنبال اولین برد
اجاره نشینی فخرمان جهان و پارالمپیک در خانه 50 متری؛ همین قدر تلخ و باورنکردنی
برومند: یامگا هم می‌تواند جباری باشد هم تیم/ استقلال به بازیکنی مثل او نیاز داشت
برترین مهاجم فوتبال با بهترین میانگین گلزنی/ مسی سوم؛ رونالدو هفتم+عکس
گفت وگو با آر پی جی زن استقلال/ از خاطرات با کلوپ و حجازی تا انتقاد از مجیدی و گلی که باعث مرگ پدرش شد+فیلم
لیگ ستارگان قطر ادربی ایرانی‌ها میان کنعانی زادگان و اسماعیلی برنده نداشت/ توقف باران ابراهیمی
میثاقیان در غیاب بازیکنان سرمربی پدیده شد/ وداع تیم مهدی با مهاجری
پرتماشاکرترین دیدارهای فوتبال پس از بازگشت هواداران به ورزشگاه/ رکورد جالب اولدترافورد+عکس
کاپیتان ذوب‌آهن: حومان باخت نبود/ شاید گل دوم استقلال آفساید بود
=====
>> Retrieval Time: --- 0.08173918724060059 seconds ---
=====
```

مدل آموزش داده شده از قبل

```
>> Please Enter your Query: ایف
>> Top 10 Results for «سینوفارم» :
=====
اعلام آمار تجمیعی واکسیناسیون کرونا در ایران تا امروز
اعلام آمار تفکیکی واکسیناسیون کرونا تا ۸ شهریور ماه
اعلام آمار تفکیکی واکسیناسیون کرونا در کشور تا ۱۴ شهریور ماه
تزریق بیش از ۳۱ میلیون دُز واکسن کرونا در کشور تا کنون
آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲۱ مرداد
اعلام آمار تفکیکی واکسیناسیون کرونا تا صبح امروز
آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲ شهریور
آمار تفکیکی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه
تزریق بیش از ۵۶ میلیون و ۵۹۶ هزار دُز واکسن کرونا در کشور تا کنون
تزریق بیش از ۴۴۷ هزار دُز واکسن در کشور در ۲۴ ساعت گذشته
=====
>> Retrieval Time: --- 0.1047506332397461 seconds ---
=====
```

مدل آموزش داده شده روی مجموعه اخبار داده شده

```
>> Please Enter your Query: ایف
>> Top 10 Results for «سینوفارم» :
=====
ورود ششمین محموله ۶ میلیون دُز "سینوفارم" به کشور
انتشار مجوز وزارت بهداشت به هلال احمر برای واردات واکسن کرونا
آیا واکسن سینوفارم برای دانش‌آموزان بی‌خطر است؟
آمار تفکیکی واکسیناسیون کرونا در کشور تا ۲۱ مرداد
آمار تفکیکی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه
اعلام آمار تفکیکی واکسیناسیون کرونا تا صبح امروز
تداوم واردات واکسن از سوی هلال احمر/ ۶ میلیون دُز سینوفارم وارد کشور شد
ورود دومین محموله سنگین واکسن کرونا به کشور از سوی هلال احمر
اعلام آمار تجمیعی واکسیناسیون کرونا در ایران تا امروز
تزریق دوز دوم واکسن کرونا در استان مرکزی طی هفته جاری
=====
>> Retrieval Time: --- 0.6022026538848877 seconds ---
=====
```

مدل tf-idf

عملکرد رضایت‌بخش‌تری را شاهدیم چرا که نتایج ارتباط بسیار بیش‌تری به کوئری دارند. علت آن می‌تواند این باشد که کلمه کوئری دارای idf بالایی بوده که در رتبه‌بندی نتایج تاثیر زیادی می‌گذارد.

در اینجا، پرسمان داده شده در دو حالت مختلف مدل word2vec، عملکردهای کاملاً متفاوتی دارد؛ به طوری که نتایج بازیابی شده برای مدل آموزش داده شده از قبل به شدت ضعیف است و نتایج کاملاً نامربوطند. احتمال قوی آن است که این کلمه در داده‌های آموزشی این مدل وجود نداشته و به همین علت هم برای مدل ناشناخته است و نتایج خوبی بدست نیامده‌اند. در مدل آموزش داده شده بر روی اخباری که داشتیم، چون این کلمه در این مجموعه داده از قبل وجود داشته بود، نتایج نسبتاً خوبی را شاهد هستیم به طوری که همگی نتایج اگر هم شامل کلمه پرسمان نباشند، به طور معنایی با آن قرابت دارند (مثلاً به دیگر واکسن‌ها مربوطند). در مدل tf-idf هم نسبت به مدل‌های آموزش دیده،



## پ) یک پرسمان دشوار و کم تکرار چند کلمه‌ای

➤ پرسمان: تفریغ بودجه

```
>> Please Enter your Query: تفریغ بودجه
>> Top 10 Results for «تفریغ بودجه» :
=====
آغاز جلسه غیرعلنی مجلس برای بررسی گزارش تفریغ بودجه ۹۹
قالبیاف: تا سه ماه آینده گزارش نهایی تفریغ بودجه ۹۹ در محن قرائت خواهد شد
سی‌ودومین گزارش تفریغ بودجه ۲۰ مهر توسط بذریاش در مجلس قرائت می‌شود
بذریاش اعلام کرد: رکوردشکنی در تفریغ بودجه ۹۹/ نظارت دیوان محاسبات بر اجرای بودجه در مقاطع ۲ ماهه
نظارت دیوان محاسبات بر اجرای بودجه سالانه در مقاطع ۲ ماهه را باید به فال نیک گرفت
تقدیر رئیس‌جمهور از اقدام دیوان محاسبات در گزارش تفریغ بودجه سال ۹۹
گام بلند دیوان محاسبات برای آنلاین شدن «تفریغ بودجه»/ پایانی بر گزارش‌های تاریخ گذشته از دخل و خرج‌ها
نشست غیرعلنی مجلس برای بررسی گزارش رئیس سازمان برنامه و بودجه
جزئیات نشست غیرعلنی مجلس درباره بودجه سال قبل/ 61 درصد منابع صندوق توسعه در دولت روحانی صرف امور جاری شد
گزارش تفریغ بودجه ۹۹ به تفکیک دستگاهی و استانی به زودی به مجلس می‌رسد
=====
>> Retrieval Time: --- 0.10913372039794922 seconds ---
=====
```

مدل آموزش داده شده از قبل

در این قسمت هم از  
آنجایی که کلمات idf  
نسبتاً بالایی دارند، در نتایج  
مربوط به مدل tf-idf

وضعیت بهتری را شاهد  
هستیم. البته نتایج مربوط  
به مدل word2vec، چه  
آموزش دیده از قبل و چه  
آموزش دیده با اخبار داده  
شده، نتایج مرتبطی هستند  
که به پرسمان نیز مربوطند  
اما می‌توان دید که احتمالاً  
Average Precision  
کمتری خواهند داشت.

```
>> Please Enter your Query: تفریغ بودجه
>> Top 10 Results for «تفریغ بودجه» :
=====
آغاز جلسه غیرعلنی مجلس برای بررسی گزارش تفریغ بودجه ۹۹
سی‌ودومین گزارش تفریغ بودجه ۲۰ مهر توسط بذریاش در مجلس قرائت می‌شود
بذریاش اعلام کرد: رکوردشکنی در تفریغ بودجه ۹۹/ نظارت دیوان محاسبات بر اجرای بودجه در مقاطع ۲ ماهه
نظارت دیوان محاسبات بر اجرای بودجه سالانه در مقاطع ۲ ماهه را باید به فال نیک گرفت
قالبیاف: تا سه ماه آینده گزارش نهایی تفریغ بودجه ۹۹ در محن قرائت خواهد شد
گام بلند دیوان محاسبات برای آنلاین شدن «تفریغ بودجه»/ پایانی بر گزارش‌های تاریخ گذشته از دخل و خرج‌ها
دولت در شش ماهه نخست امسال 115 هزار میلیارد تومان برای هزینه‌های جاری اوراق فروخت
دولت دوازدهم بارانه نقدی سه دهک بالای درآمدی را قطع نکرد
تهیه دومین گزارش نظارت مستمر ۲ ماهه بر اجرای بودجه سال ۱۴۰۰
گزارش تفریغ بودجه ۹۹ به تفکیک دستگاهی و استانی به زودی به مجلس می‌رسد
=====
>> Retrieval Time: --- 0.20602869987487793 seconds ---
=====
```

مدل آموزش داده شده روی مجموعه اخبار داده شده

```
>> Please Enter your Query: تفریغ بودجه
>> Top 10 Results for «تفریغ بودجه» :
=====
قالبیاف: تا سه ماه آینده گزارش نهایی تفریغ بودجه ۹۹ در محن قرائت خواهد شد
آغاز جلسه غیرعلنی مجلس برای بررسی گزارش تفریغ بودجه ۹۹
گزارش تفریغ بودجه ۹۹ به تفکیک دستگاهی و استانی به زودی به مجلس می‌رسد
سی‌ودومین گزارش تفریغ بودجه ۲۰ مهر توسط بذریاش در مجلس قرائت می‌شود
نظارت دیوان محاسبات بر اجرای بودجه سالانه در مقاطع ۲ ماهه را باید به فال نیک گرفت
قالبیاف: کمیسیون‌های مجلس در یک ماه نظر خود را درباره تفریغ بودجه ۹۹ اعلام کنند
بذریاش اعلام کرد: رکوردشکنی در تفریغ بودجه ۹۹/ نظارت دیوان محاسبات بر اجرای بودجه در مقاطع ۲ ماهه
نشست غیرعلنی مجلس با امیر عبداللّه‌یان، بذریاش و میرکاظمی
تقدیر رئیس‌جمهور از اقدام دیوان محاسبات در گزارش تفریغ بودجه سال ۹۹
دیدار وزیر اقتصاد و رئیس دیوان محاسبات/ بذریاش: تفریغ بودجه 1400 هم زودتر از موعد ارائه می‌شود
=====
>> Retrieval Time: --- 0.9822220802307129 seconds ---
=====
```

مدل tf-idf

