

به نام خدا

۹۷۲۳۱۰۰	محمد مهدی هجرتی
فاز دوم	پروژه بازیابی اطلاعات
آذر ۱۴۰۰	استاد نیک آبادی

در این مرحله با کمک positional index ساخته شده در فاز قبل، برای هر ترم champion list آن را شامل ۱۰۰ داک برتر ساختیم و سپس با کمک index elimination و حذف امتیاز های صفر، با استفاده از فرمول های داده شده cosine score را محاسبه کردیم و در نهایت k سند با بیشترین امتیاز را به عنوان خروجی منتشر کردیم.

بررسی کوثری ها

الف) بین الملل

```
Enter your query: اللهم لا اله الا انت
1 ) docID: 624 | title: دش یناهج یرک بحاص یرکسا
2 ) docID: 1008 | title: دندش بومنم لالقتسا هاگشرب یراس یا هفرح هتیم ریدم و للملانیب نواعم
3 ) docID: 7304 | title: یرسانش هغیش عجم و گنهرف و ملع هاگشرب یرکسا همانمهافت داقعنا
4 ) docID: 636 | title: تسا رابکی لاس 2 ره لابتوق یناهج ماج یرازگرب فلاحم کیبملا للملانیب هتیم ارچ
5 ) docID: 2779 | title: درک یقوق نیام دنوشیم رضاح یتسینوی هس میزر لباقم هک یناراکشرو زا دیاب: یسودق یمیرک
6 ) docID: 2700 | title: میوش للملانیب یدایم رد ناشقوق غیمت زا عنام نامناراکشرو را کیتاملید نیامح اب دیاب: یرنوک
7 ) docID: 5133 | title: ناریا دامیتقا یارب یناهج دیدج درواتسید کی
8 ) docID: 7386 | title: میریگب یدج ار نشیمینا یارب اونجم دولوت شروم
9 ) docID: 5975 | title: رمح لاله یوس زا روشک هب انورک نسکاو نیگنس هلومجم نیمود دورو
10 ) docID: 4669 | title: میتسه یتسود هطبار تیوقت لابند هب رزخ هیشاح یاهروشک اب طابترا رد
```

در تمام اسناد کلمه ی بین الملل یا مشتقات آن به کار رفته است و نتایج خروجی طوری مرتب شده اند که در اسناد بالاتر تعداد تکرار این ترم بیشتر از سایرین است.

ب) دانشگاه صنعتی امیرکبیر

```
Enter your query: ریکیریما یتعنس هاگشرب
1 ) docID: 2792 | title: تسا سدقم عافد همرع ام یاههگشرب طیحم زورم
2 ) docID: 2793 | title: دریگ تروص یتافیقح راک سدقم عافد نارود رد نایههگشرب شقن تبت یارب دیاب
3 ) docID: 7520 | title: یکناباب دیعس هب یرنه کی هجرد یهاوگ یاطعا
4 ) docID: 7236 | title: ییاجر رورسدمجم یارب یدوبدای یرازگرب
5 ) docID: 1959 | title: روهج سیئر لوا نواعم هب نارته یاههگشرب ییوجشناد جیپ 8 همان
6 ) docID: 5022 | title: مانبت تلهم ددم دیدم/دش حالما مراجه راب یارب اههگشرب یمادختسا نومرا یامنه ار هجرتفد
7 ) docID: 5021 | title: مانبت تلهم ددم دیدم/دش حالما مراجه راب یارب اههگشرب یمادختسا نومرا یامنه ار هجرتفد
8 ) docID: 7004 | title: ریوامت + دیبا هناع هب ییاجر رورسدمجم هقررب
9 ) docID: 2131 | title: ناسنایفا رد تیاج یلما نارمقم دوعس لآ و اکیرمآ /ناسنایفا یرگولوسنک لباقم رد رودنق دجسم یادهش تشادگررب
10 ) docID: 5089 | title: نارته ناگندیام یدیزاب رد سدق ناسرهش یبأ هروح تالکشم یررب
Enter your query:
```

در این نتیجه لیست داک ها را براساس اهمیت و تعداد تکرار کلمات کوثری مشاهده میکنیم. مثلا در نتیجه اول با اینکه ترکیب دانشگاه صنعتی امیرکبیر پشت سر هم نیامده است اما لا توجه به تکرار نسبتا زیاد ترم های دانشگاه و امیرکبیر در رنک بالایی قرار گرفته است.

اما در تصویر زیر که مربوط به جست و جو در فاز یک می باشد می بینیم که تمام اسناد صرفا بر اساس وجود یا عدم وجود کلمات کوثری لیست شده اند و در آن ها تکرار و اهمیت ترم مطرح نمی باشد.

```
Enter your query: ریکیما یثعنص هاگشناد
1 ) docID: 362 | title: ییاد هن دنروایب دای هب ناسنا ناوئع هب ارم مدرم مه اوخیم: ییاد
2 ) docID: 390 | title: ریوامت +روآلادم ناروگیئتشک و کیبملا نا نامرهق زا دارآ هاگشناد لیلت
3 ) docID: 393 | title: دوب هدرک ریغت یسور فیرح کبس/دنه دیم نم هب ار یناهج لانیف رایتما مدرکیم رکف: ینادزی
4 ) docID: 552 | title: دش میخشم ناقباسم غورش نامز /دندش یفرعم یتشک گیل نا ناابزیم
5 ) docID: 560 | title: ؟یبرم ای تسا سردم زدن انرف/روسفورپ زا اهیلاب دنه بیجع داقتنا
6 ) docID: 724 | title: ناابزیم بااختنا یارب هسلج یرازگرب/دش مالعا یتشک گیل زاغآ نامز
7 ) docID: 775 | title: تسین NBA ناقباسم هب دنم هقالع نادنج هک یتسپلانتکسپ یوناب
8 ) docID: 978 | title: نهک هراق رد تسکش ات یزاس هموزر یارب اکیرمآ هب رفس زا /نولسراب رد شزرو ریزو یقاتا مه اب
9 ) docID: 1007 | title: PCR تست یارب یدرأیلوم 2 هنیزه/دارآ یتشک گیل رد میت ۱۴ و یگنرف گیل رد میت ۱۶ روضح
10 ) docID: 1146 | title: ؟دش دهاوخ رازگرب هنوگج لابیلاو رترب گیل
Enter your query:
```

پ) ژیمناستیک

```
Enter your query: کیتسانمیژ
1 ) docID: 3615 | title: دش میخشم کیتسانمیژ نویساردف عمحم ریبد
2 ) docID: 1367 | title: یقالخاریغ تامادقا و طلئخم یاهنلااس موصوخ رد نارته کیتسانمیژ تاهه رادشه
3 ) docID: 632 | title: دش رتشیب اهتیقفوم نویساردف توکیاب اب /دنتسه کیتسانمیژ ندرک جلف لابند هب یخرب: هاوخرخ
4 ) docID: 3664 | title: یماسا + کیتسانمیژ نویساردف تسپاری تسپ یارب دزمان ۱۳ مان تبث
5 ) docID: 1456 | title: شزرو ترازو طلئخم شزرو یاهنلااس اب دروخرب و رکذت اینزتوس
6 ) docID: 4188 | title: ؟تسا عونمم نارته رد شزرو یاهتیلاعف مادک /۱۴۰۰ رهم ۹ ات ناریا شزرو یاهلیطعت تاییزج
7 ) docID: 3878 | title: ریوصت +هامرت ناآپ ات ناریا شزرو یلیطعت تاییزج
8 ) docID: 4056 | title: !درک لمع یباختنا هخرج سراسا رب الماک اب /تسا ۱۰۰ راکتسرد ،مشاب ۱۰ ین نجابم رد نم رگا: ریبد
9 ) docID: 7561 | title: ییاقشوق یاهیلال طفاح تشذگرد یارب داشرا ریزو نیلسپت مایپ
10 ) docID: 7560 | title: دوشیم شخب هدنز «امش اب هعمج حبص»
Enter your query:
```

در تمام اسناد کلمه ی ژیمناستیک به کار رفته است و نتایج خروجی طوری مرتب شده اند که در اسناد بالاتر تعداد تکرار این ترم بیشتر از سایرین است و یا آن داک کوتاه تر است.

ت) واکسن آسترازنکا

```
Enter your query: اکنزارتسآ نسکاو
1 ) docID: 5936 | title: زورما حبص ات انورک نویسانیسکاو یکیکت رامآ مالعا
2 ) docID: 5896 | title: هامدادرم ۲۶ ات ناریا رد انورک نویسانیسکاو یکیکت رامآ
3 ) docID: 5889 | title: هامدادرم ۲۶ ات ناریا رد انورک نویسانیسکاو یکیکت رامآ
4 ) docID: 5864 | title: دادرم ۲۱ ات روشک رد انورک نویسانیسکاو یکیکت رامآ
5 ) docID: 5853 | title: دادرم ۲۱ ات روشک رد انورک نویسانیسکاو یکیکت رامآ
6 ) docID: 5569 | title: دش روشک دراو انورک نسکاو یزود نویلیم ۱۰۴ هلومحم
7 ) docID: 4931 | title: دش روشک دراو انورک نسکاو یزود نویلیم ۱۰۴ هلومحم
8 ) docID: 6082 | title: زورما ات ناریا رد انورک نویسانیسکاو یعیمجت رامآ مالعا
9 ) docID: 6072 | title: زورما ات ناریا رد انورک نویسانیسکاو یعیمجت رامآ مالعا
10 ) docID: 5913 | title: رویرهش ۲ ات روشک رد انورک نویسانیسکاو یکیکت رامآ
Enter your query:
```

در اینجا نتیجه ی اول شامل دقیقا عبارت واکسن آسترازنکا نیست، اما تعداد تکرار کلمه ی آسترازنکا به علت وزن بالای آن باعث بالاتر قرار گرفتن آن شده است. در سایر اسناد که راجع به آمار واکسیناسیون می باشد نیز همین قضیه برقرار است.

اما مثلا در نتیجه ۶ میبینیم که ۱ بار کلمه ی آسترازنکا آمده ولی تعداد کلمه ی واکسن بیشتر از ۱ مورد است. اما به علت اینکه ارزش کلمه ی آسترازنکا بیشتر از واکسن می باشد این سند امتیاز زیادی دریافت نکرده است.

اما در تصویر زیر که مربوط به نتیجه حاصل از سرچ فاز یک می باشد، میبینیم که نتیجه اول دقیقا کلمه واکسن آسترازنکا آمده است و در سایر نتایج کلمه ی واکسن و یا آسترازنکا فارغ از اهمیت و تعداد آن.

Enter your query:	واکسن آسترازنکا
1) docID: 6336	title: ؟تسا ردقچ اکن زارتسا نسکاو رد نوخ هتخل داجیا ناکم
2) docID: 4931	title: دش روسک دراو انورک نسکاو یزود نویلیم ۱۰۴ هلومحم
3) docID: 5569	title: دش روسک دراو انورک نسکاو یزود نویلیم ۱۰۴ هلومحم
4) docID: 5685	title: انورک اب هزرا بم خالس نیرتمهم
5) docID: 5823	title: نویسانیسکاو یمومع هخرج زا جراخ نسکاو قیرزت زا زیهرپ/انورک اب هلباقم هار نیرتینالقع:نویسانیسکاو
6) docID: 5825	title: مینادب انورک نویسانیسکاو دروم رد دیاب هک یتاکن
7) docID: 5831	title: نویسانیسکاو یمومع هخرج زا جراخ نسکاو قیرزت زا زیهرپ/انورک اب هلباقم هار نیرتینالقع:نویسانیسکاو
8) docID: 5833	title: مینادب انورک نویسانیسکاو دروم رد دیاب هک یتاکن
9) docID: 5845	title: ناتسنمرا و ناریا ینیمز رفس یارب هزات تاررقم
10) docID: 5857	title: ؟دنراد لخدت ییاهوراد هچ اب انورک یاهنسکاو