

گزارش پروژه

سه‌د نوعی - ۹۹۲۳۰۸۷

مقدمه: در این پروژه، پس از پیش‌پردازش اولیه تمامی متون و ساخت positional index از آنها، چند نوع کلاس مختلف برای پردازش query طراحی شده است که در ادامه به ترتیب توضیح می‌دهیم.

پیش‌پردازش: برای پیش‌پردازش متون، از 3 ماژول اصلی در پوشه preprocessor استفاده شده است:

- ماژول `utils.py`: در این ماژول، ابتدا یک سری متغیرهای `global` تعریف شده است که در سرتاسر روند نرم‌الایز کردن استفاده می‌شود.
- کاراکترهای `'\u200c'` و `' '` به ترتیب به عنوان نیم‌فاصله و فاصله تکی (عادی) استفاده شده اند. برای دسترسی راحت تر، در دو متغیر `ZERO_WIDTH` و `SPACE` نگهداری شده اند.
- دیکشنری `REPLACE_MAP` برای نگهداری یک `mapping` از کاراکترهای احتمالی در داده‌ها به نماینده آن کاراکتر در پردازش‌های ما، که شامل مواردی مثل تبدیل "ی" به "ی"، تبدیل انواع مختلف `white space` ها به فاصله تکی یا نیم‌فاصله، تبدیل کاراکترهای خاص مانند "□" به نماینده آن یعنی "بسم-الله-الرحمن-الرحیم" و ...
- لیست `DELETE_CHARACTER` شامل کاراکترهایی است که باید از متن‌ها حذف شوند تا پردازش آسان‌تر شود. مانند انواع `punctuation` ها، فتحه، ضمه، حروف عربی‌ای که روی حروف قرار می‌گیرند (مانند ^ط) و ...
- لیست‌های `PRE_SUBSTRING` و `POST_SUBSTRING` یک سری زیرکلمه‌های خاص هستند که معمولاً وقتی بعد یا قبل از یک کلمه قرار می‌گیرند، باید بین آنها نیم‌فاصله باشد. از این دو لیست برای اصلاح فاصله‌های اشتباه در ادامه استفاده می‌کنیم. مثلاً زیرکلمه "وری"، بعد از یک کلمه می‌آید و باید بین آنها نیم‌فاصله باشد، مانند بهره‌وری
- لیست `PERSIAN_NUMBERS` لیست اعداد فارسی است که در ادامه استفاده می‌کنیم

ماژول `normalizer.py`:

- تابع `delete_invalid_character` از دیکشنری `DELETE_CHARACTER` استفاده می‌کند و به ازای هر رشته ورودی به این تابع، به صورت کاراکتر به کاراکتر روی آن پیمایش می‌کند، و در صورت نیاز کاراکتری را حذف می‌کند.
- تابع `replace_invalid_character` از دیکشنری `REPLACE_MAP` استفاده می‌کند و به ازای هر رشته ورودی به این تابع، به صورت کاراکتر به کاراکتر روی آن پیمایش می‌کند، و در صورت نیاز کاراکتری را جایگزین می‌کند.
- تابع `correct_number_spacing` برای ایجاد فاصله تکی بین کلمات و اعداد استفاده شده است. قبل از انجام این کار، در دیکشنری موتور جستجوی ما، کلماتی مانند "ماده 74" دیده

- می‌شد و فاصله مناسب بین عدد 74 و کلمه ماده نبود. در این تابع بین کلماتی مانند این مورد، یک فاصله تکی اضافه می‌شود.
- تابع `correct_word_spacing` برای تبدیل فاصله‌های نامناسب بین کلماتی مانند "می‌روند" و تبدیل آن به "می‌روند" استفاده شده است.
- تابع `normalize` توابع بالا را به ترتیب صدا می‌زند و خروجی هر یک را به دیگری می‌دهد و در نهایت متن `normalize` شده را به عنوان خروجی برمی‌گرداند.

- ماژول `tokenizer.py`: در این ماژول، فقط یک تابع ساده به اسم `tokenize` وجود دارد که هر رشته ورودی را با کاراکتر فاصله تکی (" ") توکنایز می‌کند و آرایه توکن‌ها را برمی‌گرداند.

- ماژول `preprocessor.py`: این ماژول، به ازای هر رشته ورودی، عملیات `normalize`, `tokenize`, `stem`, `lemmatize` را انجام می‌دهد. در رابطه با `stem` و `lemmatize`. از توابع آماده کتابخانه `hazm` برای ریشه‌یابی استفاده شده است. در این مرحله ابتدا `stem` انجام شده است. یعنی کلماتی مانند "کتاب‌ها" و "کتابی" و "کتاب‌هایی" همه به "کتاب" تبدیل می‌شوند. سپس، کلمات حاصل از این مرحله، `lemmatize` می‌شوند. یعنی حات‌های مختلف فعل‌ها را هم به یک حالت یکتا برمی‌گرداند. به عنوان مثال، دو فعل "گفته‌ام" و "گفت"، هر دو به "گفت" تبدیل می‌شوند، این کاری بود که تابع `stem` به تنهایی انجام نمی‌داد (`stem` این دو کلمه را به ترتیب به "گفته" و "گف" تبدیل می‌کند). در واقع هدف ما این بوده است که حالات مختلف یک کلمه، همه به یک کلمه در دیکشنری نهایی تبدیل شوند.

ساخت شاخص‌ها: یک ماژول اصلی در پوشه `indexConstruction` وجود دارد و یک ماژول که از آن استفاده می‌کند.

- ماژول `invertedIndex.py`: در این ماژول یک سری کلاس برای ساخت ایندکس‌ها تعریف شده است:
 - کلاس `Posting` که نماینده همان مفهوم `posting` در بازیابی اطلاعات است و شناسه سند و تعداد تکرار آن کلمه در آن سند را نگه می‌دارد.
 - کلاس `PositionalPosting` که از `Posting` ارث‌بری می‌کند و یک فیلد اضافه که از نوع لیست است را برای نگهداری مکان‌های وقوع آن کلمه در سند استفاده شده است.
 - کلاس `PostingsList` برای نگهداری تمام `Posting` های مربوط به یک کلمه استفاده می‌شود. همچنین مقادیر فرکانس بین سندها (کلمه در چند سند وجود دارد) و تعداد تکرار کل آن کلمه در

مجموعه اسناد را نگهداری می‌کند. متد مهم این کلاس، `getDocIdTFPairs` می‌باشد که یک لیست از دوتایی‌های شناسه سند و فرکانس آن کلمه در آن سند (`docId`, `tf`) را برمی‌گرداند. کلاس `InvertedIndexType` که یک `enum` است، برای تنظیم انواع ممکن برای یک `InvertedIndex` استفاده می‌شود.

کلاس `InvertedIndex` که نماینده کل شاخص‌های ماست. یک دیکشنری برای نگهداری یک `mapping` از کلمه به `PostingsList` آن دارد. همچنین یک فیلد برای نگهداری تعداد کل سندها. تابع `addPosting` به ازای کلمه و شناسه سند داده شده و فرکانس کلمه در آن سند، یک شیء از نوع `Posting` می‌سازد و به انتهای `PostingsList` آن کلمه اضافه می‌کند. تابع `deleteMostRepeated` به ازای هر کلمه، تعداد تکرار کل آن در مجموعه اسناد را می‌شمارد و با استفاده از ساختمان داده `heap`، تعداد `k` بیشترین تکرار شده را از شاخص‌ها به طور کامل حذف می‌کند. لیست کلمات حذف شده به ازای شاخص عادی و شاخص `champion` در دو فایل با نام‌های `"12k-ii-deleted-terms.pkl"` و `"12k-champII-deleted-terms.pkl"` ذخیره شده است، اما اینجا نیز هرکدام را لیست می‌کنیم.

نکته 1: برای سادگی کدزنی و استفاده مجدد از توابع پیاده‌سازی شده قبلی، برای پیاده‌سازی `efficient search` و استفاده از ایده `champion list`، به جای اضافه کردن یک لیست جدا به شاخص فعلی، یک شاخص دیگر ساخته شده است که به ازای هر کلمه، در صورتی که در یک سند حداقل 3 بار آن کلمه تکرار شده باشد، `Posting` مربوط به این سند، به آن کلمه اضافه شده است. این در نهایت باعث شده است که تعداد تکرار هر کلمه در کل اسناد در شاخص `champion` با تعداد تکرار کلمات در شاخص عادی، متفاوت باشد و در نهایت منجر به این شده است که در هنگام حذف 50 کلمه پرتکرار، بعضی از این کلمات متفاوت باشند یا ترتیب آن‌ها در عملیات حذف متفاوت بوده باشد. لیست کلمات حذف شده:

50 کلمه پرتکرار حذف شده از شاخص عادی: (تعداد تکرار، توکن حذف شده)

[('و', 234676), ('در', 164981), ('به', 136421), ('از', 92916), ('هست', 85488), ('این', 84203), ('که', 76244), ('کرد#کن', 75218), ('شد#شو', 71830), ('دار', 69359), ('با', 68904), ('را', 61387), ('اس', 30981), ('بر', 30697), ('داشت#دار', 29693), ('بود#است', 24841), ('داد#ده', 23916), ('یک', 23091), ('ت', 22360), ('کشور', 19727), ('ما', 18860), ('خود', 18859), ('بر', 18547), ('بایست#بای', 17501), ('باز', 15915), ('تا', 15820), ('های', 15191), ('اما', 14950), ('مرد#میر', 14571), ('گف', 14566), ('مجلس', 14440), ('اسلام', 14177), ('فارس', 13304), ('گزار', 12894), ('دو', 12863), ('پیا', 12829), ('کار', 12796), ('رییس', 12713), ('ایر', 12702), ('مل', 12572), ('سال', 12417), ('خبرگزار', 12353), ('انت', 11792), ('باشید#باش', 11776), ('دول', 11530), ('خواست#خواه', 11487), ('دا', 11411), ('بازیکن', 11223), ('ب', 11019), ('گرفت#گیر', 10051), ('اینکه',)]

50 کلمه پرتکرار حذف شده از شاخص champion: (تعداد تکرار، توکن حذف شده)

[232618, 'و'], [163220, 'در'], [134318, 'به'], [88532, 'از'], [80268, '#هست'], [79287, 'این'], [71367, 'که'], [69992, 'کرد#کن'], [66214, 'شد#شو'], [63947, 'با'], [63409, 'را'], [56093, 'اس'], [24918, 'داشت#دار'], [24686, 'بر'], [23895, 'بود#است'], [19395, 'ت'], [18884, 'داد#ده'], [18143, 'یک'], [17960, 'کشور'], [17030, 'ما'], [14674, 'بایست#بای'], [13842, 'باز'], [12917, 'خود'], [12541, 'مجلس'], [12236, 'های'], [12221, 'بر'], [11995, 'مرد#میر'], [10931, 'اما'], [10718, 'اسلام'], [10055, 'رییس'], [10049, 'تا'], [9807, 'دول'], [9728, 'ایر'], [9205, 'گف'], [9201, 'مل'], [8647, 'بازیکن'], [8631, 'سال'], [8593, 'کار'], [7939, 'دو'], [7848, 'انقلاب'], [7694, 'باشید#باش'], [7150, 'ب'], [6822, 'خواست#خواه'], [6773, 'باشگاه'], [6471, 'اینکه'], [6318, 'کن'], [6162, 'ها'], [6160, 'گرفت#گیر'], [6070, 'خوب'], [5999, 'جمهور']]

نکته 2: توکن‌های ذکر شده، بعضی اوقات کلمات کامل نیستند، به عنوان مثال توکن "ایر"، کلماتی بوده‌اند که پس از stem و lemmatize به این توکن تبدیل شده‌اند. مانند کلمه "ایران" که پس از این دو عملیات به این توکن تبدیل می‌شود.

نکته 3 (بسیار مهم): توکن‌های ذکر شده، به صورت کامل از شاخص‌ها حذف شده‌اند، پس اگر کوئری ما فقط شامل این کلمات باشد، یک لیست خالی به عنوان پاسخ جستجو برمی‌گرداند. مثلاً اگر در شاخص عادی، کلمه "ایران" یا "فارس" را جستجو کنیم، هیچ پاسخی برگردانده نمی‌شود. برای بهتر شدن این موضوع می‌توانستیم که با پیمایش بر روی پرتکرارترین توکن‌ها، فقط توکن‌هایی که کلمات بی‌ارزش مانند "در"، "از" و ... می‌باشند را حذف کنیم، و در این صورت کلماتی مانند "ایران" حذف نمی‌شدند. اما در این پروژه، طبق دستورکار، 50 کلمه پرتکرار حذف شده‌اند و توجهی به ارزش کلمات حذف شده نشده است.

- و در آخر ماژول make.py: در این ماژول، در ابتدا دادگان با کتابخانه json خوانده شده‌اند، سپس روی هر سند پیمایش شده است و قسمت content هر سند، پیش‌پردازش شده است و سپس به ازای هر سند، به ازای هر کلمه متفاوت آن، تعداد تکرار آن کلمه شمرده شده است و سپس Posting مربوط به آن ساخته شده است. در آخر، هر دو شاخص و توکن‌های حذف شده آنها ذخیره شده‌اند.

ساخت شاخص‌ها: یک ماژول اصلی در پوشه queryProcessing وجود دارد و یک ماژول که از می‌توان از آن برای تست موتور جستجو استفاده کرد.

- ماژول queryProcessor.py: در این ماژول یک سری کلاس برای پردازش کوئری‌ها تعریف شده است:

- کلاس QueryProcessor: یک کلاس پایه برای پردازش کوئری‌ها و تطابق با سندها می‌باشد. تابع اصلی آن findKRelevant_ است. در این تابع، کوئری وارد شده که به صورت یک رشته (string) می‌باشد پیش‌پردازش شده و به توکن تبدیل می‌شود، سپس، از بین این توکن‌ها، توکن‌هایی که در دیکشنری شاخص‌های ما وجود ندارد حذف شده است، سپس، وزن tf توکن‌های باقی‌مانده با فرمول زیر حساب شده است:

$$(1 + \log(f_{t,d}))$$

سپس، به ازای هر توکن، امتیاز هر سند (که همان tf-idf است) از postingsList های توکن‌های مشترک کوئری و دیکشنری حساب می‌شود. دقت شود که پس از پردازش تمام توکن‌ها امتیاز نهایی سندهای مرتبط حساب شده است و در هر اجرای حلقه روی توکن‌ها، فقط قسمتی از presentation آن سند در فضای برداری تعیین می‌شود. در انتها، پس از محاسبه بازنمایی برداری برای اسناد و کوئری، با استفاده از تابع getSimilarity، شباهت کسینوسی اسناد با کوئری محاسبه می‌شود. در انتها نیز، با استفاده از ساختمان داده heap، تعداد k بهترین اسناد (بیشترین شباهت‌ها) برگردانده می‌شوند.

تابع getTfidfList: در این تابع، به ازای هر توکن، امتیاز (tf-idf) تمام اسناد در PostingsList مربوط به آن توکن با توجه به فرمول زیر محاسبه می‌شود:

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D) = (1 + \log(f_{t,d})) \times \log\left(\frac{N}{n_t}\right)$$

تابع getSimilarity: در این تابع، به ازای دو بردار ورودی a و b، شباهت کسینوسی محاسبه می‌شود، دقت شود که از محاسبه شباهت به ازای توکن‌های غیرمشترک هم جلوگیری شده است.

- کلاس PositionalQueryProcessor(QueryProcessor): این کلاس به صورت امتیازی پیاده‌سازی شده است. این کلاس از QueryProcessor ارث‌بری می‌کند و تنها یک تابع مهم دارد به نام findKRelevant_ که در آن، به صورت پوینری، روی لیست‌ها و مکان‌های کلمات در هر سند پیمایش می‌کند و سندهایی که phrase query های مدنظر ما را دارند، برمی‌گرداند. نحوه پردازش در آن پیچیده است، اما خلاصه توضیح به این صورت است که: اگر کوئری وارد شده "دانشگاه صنعتی امیرکبیر" باشد، ابتدا سندهایی که سه کلمه "دانشگاه صنعتی امیرکبیر" را پشت سر هم دارند را در صورت وجود به لیست found_docs اضافه می‌کند، سپس، سندهایی که "دانشگاه صنعتی" و یا "صنعتی امیرکبیر" را دارند، به این لیست اضافه می‌کند. برای این عملیات از حلقه‌های تودرتو استفاده شده است ولی پردازش‌ها حداقل هستند و در هر مرحله، اگر سندی شرایط مدنظر را داشت، به لیست اضافه شده و به سراغ بررسی سندهای بعدی از آن نقطه به بعد می‌گردیم. در واقع، در این پروژه برای پاسخ به کوئری "می‌خواهم بروم"، این دو کلمه به صورت یک توکن مجزا در نظر گرفته نشده است،

بلکه مکان‌های این دو کلمه بررسی می‌شوند و در صورت وجود سند مربوطه بازگردانده می‌شود.

در انتها، پس از پیدا کردن اسناد دارای phrase query مد نظر ما (یا حداقل دارای sub phrase query)، در صورتی که تعداد اسناد پیدا شده کمتر از k باشند، باقی اسناد با استفاده از شباهت کسینوسی پیدا می‌شوند (چون در جستجوی phrase query ها، حداقل طول subphrase ها 2 در نظر گرفته شده است و لذا هر کلمه را به صورت جدا بررسی نمی‌کند). در آخر، یک لیست از سندهای پیدا شده بازگردانده می‌شوند.

- کلاس EfficientQueryProcessor(QueryProcessor): این کلاس از QueryProcessor ارث‌بری می‌کند و تنها کاری که می‌کند این است که در ابتدا جستجو را در شاخص champion انجام می‌دهد، و در صورتی که تعداد نتایج کمتر از k بودند، جستجو را روی شاخص عادی انجام می‌دهد و نتایج را برمی‌گرداند.

- ماژول test.py: این ماژول تنها برای تست موتور جستجو نوشته شده است و در آن، پس از لود کردن دادگان و شاخص‌های عادی و champion، سه نوع کلاس از انواع QueryProcessor ها می‌سازد و نتایج جستجوی هر کدام از آنها را به ازای کوئری داده شده، پرینت می‌کند. **توجه:** در نتایج پرینت شده به ازای هر جستجو، مدت زمان پاسخگویی به آن کوئری هم پرینت می‌شود، و مدت زمان بسیار کمی است ولی دلیل اجرای کند ماژول test.py، لود کردن دادگان و شاخص‌ها از روی دیسک می‌باشد.

پاسخ به پرسش‌های مختلف: پرسش‌های مختلف با QueryProcessor های مختلف را تست و تشریح می‌کنیم. **دقت شود** که این شاخص‌ها، نمی‌توانند به کوئری‌هایی که تمام کلمات آن از کلمات حذف شده شاخص‌هاست پاسخ دهند، و باید حداقل یک کلمه در کوئری، در دیکشنری شاخص وجود داشته باشد. توجه شود که برای همه پرسش‌ها $k = 10$ سند برتر برگردانده شده است.

(الف) کلمه ساده و متداول تک کلمه‌ای: اخبار

- با استفاده از شاخص عادی: (زمان صرف شده = 0.0)

```

normal cosine search spent time: 0.0, result is:
DocID in json: 48, Score: 1.0, Title: /دیمر اپیاس هب دنول امک ، دآپ آ مرخ رد یس اسج ا یزاب [کی گیل مود و تسویب هتفه URL: https://www.farsnews.ir/news/14001223000874/دیمر-اپیاس-هب-دنول-امک-د-آ-م-ر-خ-رد-یس-اسج-ا-ی-ز-اب-کی-گیل-م-ود-و-تسویب-هتفه
****

DocID in json: 78, Score: 1.0, Title: دش نمرژ نس تیاده هنیژگ هنوئیم یسهراب رد ین ارجب تیغیو یسهراب URL: https://www.farsnews.ir/news/14001223000537/دش-نمرژ-نس-تیاده-هنیژگ-هنوئیم-یسهراب-رد-ین-ارجب-تیغیو-یسهراب
****

DocID in json: 79, Score: 1.0, Title: دندراد سن اش یلع دیس رانک رد هنیژگ 2 /دش منخشم یبرد رواد یفرعم نامز URL: https://www.farsnews.ir/news/14001223000522/دندراد-سن-اش-یلع-دیس-رانک-رد-هنیژگ-2-دش-منخشم-یبرد-رواد-یفرعم-نامز
****

DocID in json: 142, Score: 1.0, Title: ی یی اپورا داوس نایم شاه /دننک راذگا و نم هب ار یلم میت تیاده دنن اوتیم هک دناه دش هجوتم مدرم چچوکسا URL: https://www.farsnews.ir/news/14001222000329/ی-یی-اپورا-داوس-نایم-شاه-دننک-راذگا-و-نم-هب-ار-یلم-میت-تیاده-دنن-اوتیم-هک-دناه-دش-هجوتم-مدرم-چچوکسا
****

DocID in json: 148, Score: 1.0, Title: رد یعیط یارب 13 یسجن /سهل و پسرب یراودیم ا و هیشاح ینشاج اب لالقتسا درب [رترب گیل 22 هتفه زا هتکن 8 URL: https://www.farsnews.ir/news/14001221000441/رد-یعیط-یارب-13-یسجن-سهل-و-پسرب-یراودیم-ا-و-هیشاح-ینشاج-اب-لالقتسا-درب-رترب-گیل-22-هتفه-زا-هتکن-8
****

DocID in json: 203, Score: 1.0, Title: دالوف نزلگ ندرکن یل احشوخ و یرابخا تیغیو [دالوف و روتکارت یزاب هیشاح URL: https://www.farsnews.ir/news/14001221000752/دالوف-نزلگ-ندرکن-یل-احشوخ-و-یرابخا-تیغیو-دالوف-و-روتکارت-یزاب-هیشاح
****

DocID in json: 211, Score: 1.0, Title: ف-رامین /تفرگ میمیت شا هدی آ یارب یسم لثم رامین URL: https://www.farsnews.ir/news/14001221000592/ف-رامین-تفرگ-میمیت-شا-هدی-آ-یارب-یسم-لثم-رامین
****

DocID in json: 216, Score: 1.0, Title: دتی انویرتسجنم یسهلگن ا هراتس یقی ایران هب کینگن ار عطاق خس اپ URL: https://www.farsnews.ir/news/14001221000445/دتی-انویرتسجنم-یسهلگن-ا-هراتس-یقی-ایران-هب-کینگن-ار-عطاق-خس-اپ
****

DocID in json: 314, Score: 1.0, Title: پاستا و تچ «کلی زرب هراتس / نکتخر رد امورانود اب اوعد زا رامین یرگاشفا URL: https://www.farsnews.ir/news/14001220000139/پاستا-و-تچ-کلی-زرب-هراتس-نکتخر-رد-امورانود-اب-اوعد-زا-رامین-یرگاشفا
****

DocID in json: 482, Score: 1.0, Title: دتسا هوشن ییاهن یبرد نامز ریغت یدهیم /روشک ییاهه اگشزرو مامت رد ناراد اوه روضح طورش URL: https://www.farsnews.ir/news/14001217000675/دتسا-هوشن-ییاهن-یبرد-نامز-ریغت-یدهیم-روشک-ییاهه-اگشزرو-مامت-رد-ناراد-اوه-روضح-طورش
****

```

آیدی و امتیاز سندهای بازگردانده شده: (به ترتیب، 10 سند برتر)

48 -> 1.0
78 -> 1.0
79 -> 1.0
142 -> 1.0
148 -> 1.0
203 -> 1.0
211 -> 1.0
216 -> 1.0
314 -> 1.0
482 -> 1.0

به عنوان مثال، محتوای سند با آیدی 48 به شرح زیر است:

"n\به گزارش خبرنگار ورزشی خبرگزاری فارس، دیدارهای هفته بیست و دوم لیگ یک در آخرین چهارشنبه سال برگزار می‌شود و هفته بیست و سوم این رقابت‌ها 10 فروردین 1401 برگزار خواهد شد. ملوان تیم بدون شکست لیگ همچنان صدرنشین است و بعد از تساوی خانگی مقابل رایکا باید این هفته به مصاف پارس جنوبی جم برود. مسابقه‌ای که می‌تواند بسیار جذاب و دیدنی شود. پارس جنوبی رده پنجمی در این هفته کار سختی مقابل صدرنشین مقتدر لیگ دارد. قطعا سایر تیم‌های بالانشین لیگ امیدوارند پارس جنوبی از ملوان امتیاز بگیرد تا صدر جدول جذاب‌تر از قبل دنبال شود. البته اگر ملوان در این دیدار بازنده شود همچنان صدرنشین لیگ خواهد بود اما قطعا قوی سپید خزر می‌خواهد رکورد شکست ناپذیری خود را حفظ کند. احتمال دارد در این هفته دو مدافع سرباز استقلال برای ملوان به میدان بروند. سیاوش یزدانی و حسین مرادمند هنوز در لیگ یک بازی نکرده‌اند هرچند برخی رسانه‌ها مدعی شده‌اند که اصلا ملوان نیازی به جذب مدافع نداشته است. مس کرمان تیم دوم جدول لیگ این هفته بازی سختی با خوشه طلایی ساوه دارد. کرمانی‌ها 4 امتیاز کمتر از ملوان دارند و به هیچ عنوان دوست ندارند بعد از دو برد متوالی امتیاز از دست بدهند از سویی دیگر خوشه طلایی تیم رده ششمی جدول هم اگر در این بازی پیروز نشود می‌توان گفت شانسش برای صعود حتی روی کاغذ بسیار کم می‌شود. آرمان گهر سیرجان در رده سوم جدول ایستاده و میهمانی ناخوانده برای مدعیان صعود لیگ یک است. آرمان گهر حالا باید به مصاف ویستا توربین تهران برود که با 23 امتیاز در رده سیزدهم جدول لیگ ایستاده این تیم تازه وارد لیگ دسته اول می‌خواهد در میانه‌های جدول باقی بماند سقوط نکند به همین دلیل هم انگیزه‌های زیادی برای این بازی دارد. در خرم آباد شاهد یک بازی دراماتیک هستیم! جایی که خیبر خرم آباد با فراز کمالوند سرمربی جدیدش میزبان تیم سایپا تهران است. تیمی که قبل از سقوط به لیگ یک هدایتش را همین کمالوند بر عهده داشت. فراز به خیبر آمده تا بتواند خیبر رده چهارمی را به لیگ برتر بیاورد اگر فراز مقابل تیم سابق خود در سایپا پیروز نشود در آنصورت شرایط این تیم در هفته‌های آینده برای حضور در رده چهارم جدول هم سخت خواهد شد. خیبر 5 هفته است که برنده نشده اما سایپا رده هشتمی هفته گذشته هم پیروز شد و در صورت پیروزی در این مسابقه شانس صعود سه پله‌ای در جدول لیگ را دارد. شهرداری آستارا بعد از 4 برد هفته گذشته شکست دور انتظاری مقابل شمس آذر قزوین را تجربه کرد. شاگردان پایان رفت این هفته کار سختی مقابل تیم قعرنشین و بحران زده شاهین بوشهر نخواهند داشت تیمی که 4 هفته است رنگ برد را ندیده و مشکلات حاشیه‌ای و مدیریتی این باشگاه هم مدام در صدر **اخبار فوتبال** جنوب کشور است. مس شهر بابک هم با تیم شمس آذر قزوین بازی دارد. مس شهر بابک در 5 هفته آخرش در لیگ هیچ بردی نداشته و 4 مساوی و یک باخت حاصل کار این تیم بوده است. شمس آذر اما هفته گذشته نوار بردهای تیم پایان رفت را قطع کرد و ممکن است این هفته هم بتواند برای مس شهر بابک در دسرساز شود. گفته می‌شد استقلال خوزستان از مدعیان اصلی صعود است اما این تیم اهوازی تا رده دهم جدول لیگ سقوط کرده و این هفته هم باید دربی خوزستانی را با استقلال ملاتانی برگزار کند. تیمی که در رده پانزدهم است و برای بقا تلاش می‌کند. هر دو تیم هفته‌هاست در لیگ برنده نشده‌اند و مساوی هم خیلی در شرایط دو تیم در جدول لیگ تغییر ایجاد نخواهد کرد. شهرداری همدان تیم رده یازدهمی و تازه وارد لیگ یک این هفته به مصاف رایکا بابل می‌رود که در 5 هفته اخیر در لیگ فقط 5 امتیاز دشت کرده است. رایکا رده شانزدهمی هم برای بقا تلاش می‌کند. قشقایی شیراز هم با آخرین تیم جدول یعنی ماشین سازی باید روبه‌رو شود که شاید بتوان گفت دیگر امیدی برای بقا نداشته باشد. برنامه مسابقات هفته بیست و دوم لیگ یک: چهارشنبه بیست و پنجم اسفند شهرداری آستارا - شاهین شهرداری بوشهر ۱۴:۴۵ بیست و دوم وحدت آستارا گیلان مس شهر بابک - شمس آذر قزوین ۱۵:۰۰ بیست و دوم شهید سلیمانی شهر بابک کرمان خیبر خرم آباد- سایپا تهران ۱۵:۰۰ بیست و دوم تختی خرم‌آباد لرستان رایکا بابل - شهرداری همدان ۱۵:۰۰ بیست و دوم شهدا سمرق کلاکلا مازندران قشقایی شیراز - ماشین سازی تبریز ۱۵:۰۰ بیست و دوم حافظیه شیراز فارس ویستا توربین تهران - آرمان گهر سیرجان ۱۵:۰۰ بیست و دوم غدیر تهران تهران مس کرمان - خوشه طلایی ساوه ۱۵:۰۰ بیست و دوم شهید باهنر کرمان کرمان استقلال خوزستان- استقلال ملاتانی ۱۵:۳۰ بیست و دوم نفت اهواز خوزستان پارس جنوبی جم بوشهر- ملوان بندر انزلی ۱۵:۳۰ بیست و دوم تختی جم بوشهر جدول لیگ یک: 1- ملوان 47 امتیاز 2- مس کرمان 43 امتیاز 3- آرمان گهر 39 امتیاز 16- رایکا بابل 20 امتیاز 17- شاهین شهرداری بوشهر 10 امتیاز 18- ماشین سازی تبریز 6 امتیاز انتهای پیام/n\n\n"

توضیح: دقت شود که چون کوئری ما تک کلمه‌ای است، tf-idf برای آن، چون در محاسبه شباهت کسینوسی نرمال می‌شود، پس در نهایت اندازه term-frequency در این فرایند نرمال شدن از بین می‌رود، پس سندی که در آن فقط یک بار یا چندین بار، کلمه "اخبار" آمده باشد، با هم امتیاز cosine similarity یکسان خواهند داشت (score = 1.0)، پس به سندهایی که آیدی آنها کمتر بوده است را به ترتیب برگردانده است.

همچنین دقت شود که چون کوئری ما یک کلمه بود، نتایج PositionalQueryProcessor و QueryProcessor یکسان شد، ولی نتیجه EfficientQueryProcessor متفاوت است:

نتایج champion: (همه سندهای برگردانده شده، tf برای آن کلمه در آن سند، حداقل برابر با 3 بوده است، می‌توان این مورد را آزمایش کرد، اما در این گزارش آورده نشده است)

champ list + normal cosine search spent time: 0.0, result is:

DocID in json: 1322, Score: 1.0, Title: ن شپ در ات ره شاق و داب آخ زا زورف اق و يزاج ل القسا ه ايس هبع يم اهل ا اب هژيو يگوتفگ URL: https://www.farsnews.ir/news/14001206000566, مليف يونه لق داه زا-زورف-اق-و-يزاج-ل-القسا-ه-ايس-هبع-يم-اهل-ا-اب-هژيو-يگوتفگ/

DocID in json: 1428, Score: 1.0, Title: وش و سيلوپسپ كي يانودارام لگ /دشن لاجشوخ يلو درب هك يل القسا |رترب گل مه دزون هتفه زا هتكن 8 URL: https://www.farsnews.ir/news/14001204001243, ون ايس هب ايكديون ك لگ-دشن-ل-اجشوخ-يلو-درب-هك-يل-القسا-|رترب-گل-م-ه دزون-هتفه-زا-هتكن-8/

DocID in json: 1477, Score: 1.0, Title: نيرتهب يرايخ ا/مينك بذج لمف مين رد ار يبوخ نانكيزاب مئسن اوتن :ودلوس URL: https://www.farsnews.ir/news/14001204001119, نيرتهب-ي-رايخ-ا-م-ينك-بذج-لمف-مين-رد-ار-ي-بوخ-ن-انكيزاب-م-ئسن-ا-وتن-ودلوس/

DocID in json: 1890, Score: 1.0, Title: دج روتكارت /زيريت رد ينقن ردم تيبث اب يدپم نادركاش يدامتقا يادرب هم ادا |رترب گل مه دجه هتفه URL: https://www.farsnews.ir/news/14001129000631, تيبث-اب-يدي-م-ن-ادركاش-يدامتقا-ي-ادرب-هم-ادا-|رترب-گل-م-ه دجه-هتفه-

DocID in json: 1974, Score: 1.0, Title: يدرا 24 مچم /دوب مچم درخ سيئر تايه م اداق ا م ايب :يلوس |نويس اردف ن اريدم هتغ اس 9 سيلج م هم رايخ ا URL: https://www.farsnews.ir/news/14001128000761, سيئر-تايه-م-ادقا-م-ايب-يلوس-ا-نويس-اردف-ن-اريدم-هتغ-اس-9-سيلج-م-هم-رايخ-ا-را-زگرب-شه-ب-ه

DocID in json: 2315, Score: 1.0, Title: ه اكن ايسآ هيمس و لودج هب ديابن/م دركن يداش ين اكير لگ هزادن ا ناهج مشش مئس ندرپ زا دعب :نايروصنم URL: https://www.farsnews.ir/news/14001123000760, ديابن-م-دركن-يدا-اش-ين-اكير-لگ-ه-زادن-ا-ناهج-مشش-م-ئس-ن-درپ-زا-دعب-نايروصنم-

DocID in json: 2388, Score: 1.0, Title: ل اس-19-سراف يرازگريخ/دش هل اس 19 سراف يرازگريخ URL: https://www.farsnews.ir/news/14001122000809, دش-ه

DocID in json: 3781, Score: 1.0, Title: URL: https://www.farsnews.ir/news/14001104000419, زا-ليق-اهي-انورك-تايئزج-هرايرد-قارع-نويس اردف لوئسم-اب-سراف-يگوتفگ/

DocID in json: 3805, Score: 1.0, Title: تسا دايژ اههتشر نيپ يگديسر فاكش /م راد هم انرب سيراپ كيچم ل ايارب :يغورف URL: https://www.farsnews.ir/news/14001104000142, تسا-دايژ-اههتشر-نيپ-يگديسر-فاكش-م-راد-هم-انرب-سيراپ-كيچم-ل-ايارب-يغورف/

DocID in json: 4082, Score: 1.0, Title: URL: https://www.farsnews.ir/news/14001101000140, دش-در-سيلوپسپ-ه-اگشاب-رد-ديدج-ياه-بامتن-ا-دش-در-سيلوپسپ-ه-اگشاب-رد-ديدج-ياه-بامتن-ا-

ب) پرسمان ساده چند کلمه‌ای: لیگ برتر آسیا

- نتایج QueryProcessor (شاخص عادی، شباهت کسینوسی):

```

ormation retrieval/Project/queryProcessing/test.py
normal cosine search spent time: 0.009063243865966797, result is:
DocID in json: 4071, Score: 0.9999999696316004, Title: نتایج پردازش داده‌های شبکه: نتایج پردازش داده‌های شبکه, URL: http://www.farsnews.ir/news/14001101000195/
****

DocID in json: 478, Score: 0.9999997630567755, Title: پیروزی در گول-دروم / پیروزی در گول-دروم, URL: https://www.farsnews.ir/news/14001217000715/
****

DocID in json: 29, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001224000282/
****

DocID in json: 104, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001222000894/
****

DocID in json: 281, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001220000582/
****

DocID in json: 429, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001218000219/
****

DocID in json: 951, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001211000316/
****

DocID in json: 1208, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001208000394/
****

DocID in json: 1354, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001206000397/
****

DocID in json: 1597, Score: 0.999973351768449, Title: گول-دروم / گول-دروم, URL: https://www.farsnews.ir/news/14001203000518/
****

```

نتایج PositionalQueryProcessor:

```
positional + normal cosine search spent time: 0.6257209777832031, result is:
DocID in json: 1219, Score: PHRASE-QUERY-RESULT, Title: تارام ا هروطسا هب ل القتسا قباس هراتس هیئت/دش یله ال ا یرم رس زا داقتن ا شاع اب یدئاق, URL: https://www.farsnews.ir/news/14001208000297/
****

DocID in json: 6007, Score: PHRASE-QUERY-RESULT, Title: AFC: یلوپسرب درب و نه ايس شزغل م وادت/دش ناری گیل نیشتنردس ر ابرپ هتفه اب ل القتسا, URL: https://www.farsnews.ir/news/14001006000616/AFC/
****

DocID in json: 6828, Score: PHRASE-QUERY-RESULT, Title: یاهدارا مینکیم رکف /مینکیم یزاب رت ه ناراودیم ا ه دنی آ رد درب نی ا اب یدمچم لگ, URL: https://www.farsnews.ir/news/14000924000697/
****

DocID in json: 4071, Score: 0.9999999696316004, Title: دوش م امت دیاب نه ايس ینزلگ ن ارب: نی ایدی شمع, URL: https://www.farsnews.ir/news/14001101000195/
****

DocID in json: 478, Score: 0.9999997630567755, Title: یبرد /دریگب رارق افوف لاوس دروم یداز آ م وید اتسا مقاون دیا ش: رترب گیل تا قبا سم لوئسم, URL: https://www.farsnews.ir/news/14001217000715/
****

DocID in json: 29, Score: 0.999973351768449, Title: دنکیم ن تبجس رفن کی اب یه اغشاب چه دن ادم یو اقم: م وینیم ول آ لم اعریدم, URL: https://www.farsnews.ir/news/14001224000282/
****

DocID in json: 104, Score: 0.999973351768449, Title: دنک یزاب یراجتن ا ل القتسا لباقم دیاب سولوپسرب: ین ادمچم, URL: https://www.farsnews.ir/news/14001222000894/
****

DocID in json: 281, Score: 0.999973351768449, Title: ایند گیل نی رتهب زا زم وگ /میتسین یاهفرح لئاسم زا یلیخ رد: م انوکن, URL: https://www.farsnews.ir/news/14001220000582/
****

DocID in json: 429, Score: 0.999973351768449, Title: دن دشتن اهنیرت نارگ وزج یهللارون و یدئاق/تارام ا گیل یاهنیرتنزرا وزج دالوف بقر, URL: https://www.farsnews.ir/news/14001218000219/
****

DocID in json: 951, Score: 0.999973351768449, Title: یلیع امسا ناری لباقم ینوچ ام ارتسا یارب هت اباد یدیلک شقن, URL: https://www.farsnews.ir/news/14001211000316/
****
```

نتایج (EfficientQueryProcessor(champion):

```

champ list + normal cosine search spent time: 0.0, result is:
DocID in json: 777, Score: 0.9999873210472635, Title: نم انوکن /رد- ار طیارش دی اب-م انوکن/ URL: https://www.farsnews.ir/news/14001214000673/
****

DocID in json: 1193, Score: 0.9999284195444506, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14001208000657/
گ- اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14001208000657/
****

DocID in json: 337, Score: 0.9999119463353501, Title: /تسویپ ن اتسزوخد الوف هب یی این اپسا /رگنیو / URL: https://www.farsnews.ir/news/14001219000699/
تسویپ-ن-اتسزوخد-الوف-هب-یی-این-اپسا-رگنیو- URL: https://www.farsnews.ir/news/14001219000699/
****

DocID in json: 1241, Score: 0.9999119463353501, Title: /دندروآ /همورا ل ابعلا و رس رب یی الپ هج /خآولپ رد اهردرایلم رقیقت / URL: https://www.farsnews.ir/news/14001206000794/
دندروآ-همورا-ل-ابعلا-و-رس-رب-یی-الپ-هج-خآولپ-رد-اهردرایلم-رقیقت- URL: https://www.farsnews.ir/news/14001206000794/
****

DocID in json: 6755, Score: 0.9999119463353501, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14000922000193/
مدمجملگ-اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14000922000193/
****

DocID in json: 4441, Score: 0.9998811485041236, Title: /رناش /اناه اپس هورگ اب یی انشآ / URL: https://www.farsnews.ir/news/14001027000202/
رناش-اناه-اپس-هورگ-اب-یی-انشآ- URL: https://www.farsnews.ir/news/14001027000202/
****

DocID in json: 5259, Score: 0.9998017587443513, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14001017000331/
مدمجملگ-اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14001017000331/
****

DocID in json: 387, Score: 0.9997862954410672, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14001218000832/
مدمجملگ-اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14001218000832/
****

DocID in json: 163, Score: 0.999708838773586, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14001221000302/
مدمجملگ-اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14001221000302/
****

DocID in json: 1010, Score: 0.9997013698367228, Title: /مدمجملگ /اهمیت یخرب یراک غولش و راگنریخ کی یراداو له لاسو هب شنگاو / URL: https://www.farsnews.ir/news/14001210000134/
مدمجملگ-اهمیت-یخرب-یراک-غولش-و-راگنریخ-کی-یراداو-له-لاسو-هب-شنگاو- URL: https://www.farsnews.ir/news/14001210000134/
****

```

توضیح: همانطور که مشاهده می‌شود، score (در پرنیت می‌توانید مشاهده کنید) برای نتایج بازگردانده شده توسط QueryProcessor بیشتر از EfficientQueryProcessor است، اما زمان پردازش efficient کمتر از حالت دیگر بوده است. حال با هم عملکرد Positional را بررسی کنیم، اولین سند بازگردانده شده، شناسه 1219 دارد. محتوای آن: (سندی وجود نداشته که کل عبارت "لیگ برتر آسیا" در آن باشد، پس subphrase ها را برگردانده است)

"به گزارش خبرگزاری فارس، تیم فوتبال شباب الاهلی دبى شب گذشته در هفته هفدهم لیگ برتر امارات به مصاف شارجه رفت و با نتیجه یک بر صفر شکست خورد. احمد نوراللهی بازیکن ملی پوش کشورمان در این بازی برای شباب الاهلی 90 دقیقه بازی کرد اما مهدی قائدی ستاره سابق استقلال در این بازی تنها 5 دقیقه فرصت بازی پیدا کرد. در تحلیل این بازی در شبکه «ابوظبى اسپورت»، موسی عباس یکی از کارشناس اماراتی به شدت از مهدی علی سرمربی شباب الاهلی دبى به خاطر نیمکت نشین کردن مهدی قائدی انتقاد کرد و گفت: مهدی علی مربی بزرگی است و سابقه مربیگری در تیم ملی امارات را دارد. او نباید درباره مسائل داورى صحبت کند و این موضوع را باید به اهل فنش بسپارد. وظیفه کاپیتان مهدی این است تا درباره تاکتیک تیم بهترین تصمیم را بگیرد. وی افزود: عجیب ترین تصمیم در این بازی نیمکت نشین کردن بازیکنی مثل مهدی قائدی است. بدون تعارف قائدی در لیگ امارات و در هر لیگ آسیایی یکی از بهترین بازیکنان آسیا به شمار می آید. این بازیکن توانایی بالایی دارد و دلیل نیمکت نشینی اش را متوجه نمی شوم. این کارشناس در ادامه تمجید از ستاره ایرانی، گفت: مهدی قائدی بازیکنی شبیه اسماعیل مطر (اسطوره تیم ملی امارات) است. او بازیکنی با ویژگی ها و توانایی بالا است و باید آزادی عمل بیشتری داشته باشد تا بتواند تاثیرگذار باشد. او در این بازی کمتر از 5 دقیقه فرصت پیدا کرد و این برای بازیکنی در سطح قائدی کافی نیست. انتهای پیام/ n\n\n"

پ) پرسمان دشوار تک کلمه‌ای: کریسمس

نتایج QueryProcessor:

normal cosine search spent time: 0.0, result is:
DocID in json: 5431, Score: 1.0, Title: تاراغتفا+تنگژاب دیردام لائرب نادین وحقو URL: <https://www.farsnews.ir/news/14001014000983> نتقو
تاراغتفا-تنگژاب-دیردام-لائرب-نادین-ادی-زی

DocID in json: 5483, Score: 1.0, Title: ددرگیم رب سپراپ دب ین امز هج یمس URL: <https://www.farsnews.ir/news/14001014000228> ین امز-هج-یمس
ددرگیم-رب-سپراپ-دب-

DocID in json: 5921, Score: 1.0, Title: زیسرسن اتس اتم-یتسه-یزورپ هب موکجم ن ام-کلس-دجسم تنقن لباقم یم اهلا URL: <https://www.farsnews.ir/news/14001007000852> زیسرسن اتس اتم-یتسه-یزورپ-هب-م-وکجم-ن ام-کلس-دجسم-تنقن-لباقم-یم-اهلا

DocID in json: 5926, Score: 1.0, Title: سگج-دن ادرگرب-لیصحت-لحم-هب-ار-کدوک-4ن-اچن-سفر-سم-یج-راخ-مچ-اهم URL: <https://www.farsnews.ir/news/14001007000809> سگج-دن ادرگرب-لیصحت-لحم-هب-ار-کدوک-4ن-اچن-سفر-سم-یج-راخ-مچ-اهم

DocID in json: 5933, Score: 1.0, Title: سگج-یواژ هب الوی دراوگ سم سیرک هیده یماین اپسا هراتس URL: <https://www.farsnews.ir/news/14001007000739> سگج-یواژ-هب-الوی-دراوگ-سم-سیرک-هیده-یماین-اپسا-هراتس

DocID in json: 6117, Score: 1.0, Title: سم سیرک-رطاخ هب اهریم اب یل اغترب درم الفتح/خفر بقل «روتاتکی» شوریک URL: <https://www.farsnews.ir/news/14001005000165> سم سیرک-رطاخ-هب-اهریم-اب-یل اغترب-درم-الفتح-خفر-بقل-روتاتکی-شوریک

DocID in json: 6120, Score: 1.0, Title: دیدج لاس هن اتسآ رد لایتوف ه اغشزرو رد راتشک URL: <https://www.farsnews.ir/news/14001005000143> دیدج-لاس-هن اتسآ-رد-لایتوف-ه اغشزرو-رد-راتشک

DocID in json: 6148, Score: 1.0, Title: یه ار-اردس لاس اکوین-ای-ن ورترو/نومزآ بیج یارب نول هب یملگن ا میت هبرش URL: <https://www.farsnews.ir/news/14001004000708> یه ار-اردس لاس اکوین-ای-ن ورترو/نومزآ-بیج-یارب-نول-هب-یملگن ا-میت-هبرش

DocID in json: 6697, Score: 1.0, Title: یکسای ایاتن/تشاذگ ریئات یلیخ ام تئنهذ یور شوریک/دوب رتهب یملج هب نم نادرگرب تنگ یملط شوخین اهع URL: <https://www.farsnews.ir/news/14000927000153> یکسای ایاتن/تشاذگ-ریئات-یلیخ-ام-تئنهذ-یور-شوریک/دوب-رتهب-یملج-هب-نم-نادرگرب-تنگ-یملط-شوخی-ن-اهع

DocID in json: 9413, Score: 1.0, Title: یسکار-هلل اتسآ اب یئدی ا هتعلق ا ن اگدن یا من هسلج: روهمچ یسکار نواعم URL: <https://www.farsnews.ir/news/14001006000132> یسکار-هلل اتسآ-اب-یئدی-ا-هتعلق-ا-ن اگدن-یا-من-هسلج: روهمچ-یسکار-نواعم

نتائج Efficient:

```

champ list + normal cosine search spent time: 0.0, result is:
DocID in json: 6117, Score: 1.0, Title: سبک+سبک سبک رطاح هب اهیرم اب یل اغترب درم فالتخا/تفرگ بقل «روتانکده» شورک, URL: https://www.farsnews.ir/news/14001005000165/سبک سبک رطاح هب اهیرم اب یل اغترب درم فالتخا تفرگ بقل روتانکده شورک
****

DocID in json: 5431, Score: 1.0, Title: تقو/تاراحتفا+تشتگزاب دیردام لائار هب نادیز یتقو, URL: https://www.farsnews.ir/news/14001014000983/تاراحتفا+تشتگزاب دیردام لائار هب نادیز یتقو
****

DocID in json: 5483, Score: 1.0, Title: ینامز-هچ-سبک, URL: https://www.farsnews.ir/news/14001014000228/ددرگ یمرب سیراپ هب ینامز هچ سبک
ددرگ-یمرب-سیراپ-هب-
****

DocID in json: 5921, Score: 1.0, Title: درادن مه نیرمت نیمز کی ناردنزام زیسرس ناتسا/میتسه یزوریپ هب موکحم نامیلس دجسم تفتن لباقم نیم اول, URL: https://www.farsnews.ir/news/14001007000852/زیسرس-ناتسا-میتسه-یزوریپ-هب-موکحم-نامیلس-دجسم-تفتن-لباقم-نیم-اول
****

DocID in json: 5926, Score: 1.0, Title: سبک+دنادرگرب لیصحت لجم هب ار کدوک 4,ن اجن سبقر سبج راخ مچاهم, URL: https://www.farsnews.ir/news/14001007000809/سبک-دنادرگرب-لیصحت-لجم-هب-ار-کدوک-4-ن-اجن-سبقر-سبج-راخ-مچاهم
****

DocID in json: 5933, Score: 1.0, Title: سبک+یواژ هب الویدراوگ سبقرک هیده ییاین اپسا هراتس, URL: https://www.farsnews.ir/news/14001007000739/سبک-یواژ-هب-الویدراوگ-سبقرک-هیده-ییاین-اپسا-هراتس
****

DocID in json: 6120, Score: 1.0, Title: سبک+دیوچ لاس هب ناتسا رد لایتوف هکشتزو رد راتشک, URL: https://www.farsnews.ir/news/14001005000143/دیوچ-لاس-هب-ناتسا-رد-لایتوف-هکشتزو-رد-راتشک
****

DocID in json: 6148, Score: 1.0, Title: سبک+دوش یم رترب گول یه ار رادرس لاساکوین ای نوتروا/نومزآ بیج یارب نوئل هب سبقرک میت هبرش, URL: https://www.farsnews.ir/news/14001004000708/یه-ار-رادرس-لاساکوین-ای-نوتروا-نومزآ-بیج-یارب-نوئل-هب-سبقرک-میت-هبرش
****

DocID in json: 6697, Score: 1.0, Title: ککسا چیاتن/تشاوگ ریثات ولیخ ام تنهذ یور شورک/دوب رتهب سبقرک هب نم نادرگرب تفتن سبقرک/تختین اچ, URL: https://www.farsnews.ir/news/14000927000153/ککسا-چیاتن-تشاوگ-ریثات-ولیخ-ام-تنهذ-یور-شورک-دوب-رتهب-سبقرک-هب-نم-نادرگرب-تفتن-سبقرک-تختین-اچ
ام-تنهذ-
****

DocID in json: 9413, Score: 1.0, Title: دوشیم رازگرب سبقرک هلاقی آ اب ییاین اهتعلقا ن اگدن یام هسلج/روه سبقرک نواعم, URL: https://www.farsnews.ir/news/14001006000132/دوشیم-رازگرب-سبقرک-هلاقی-آ-اب-ییاین-اهتعلقا-ن-اگدن-یام-هسلج-روه-سبقرک-نواعم
****

```

توضیح: همانطور که مشاهده می‌شود، در جستجو با Efficient، ابتدا لیست champion برگردانده شده، به عنوان مثال، محتوای سند 6117 به صورت زیر است: (حداقل 3 بار کریسمس در آن تکرار شده بوده):

"n\به گزارش خبرگزاری فارس، کارلوس کی روش سرمربی نام آشنا برای ایرانی‌ها این روزها در تیم ملی فوتبال مصر حاشیه‌های زیادی دارد و موافقان و مخالفانی پیدا کرده است. در جدیدترین خبر عبدالناصر زیدان یکی از خبرنگاران مطرح ورزشی مصر به شدت به کارلوس کی روش هجوم آورد او را «دیکتاتور» نامید. این خبرنگار در مصاحبه با سایت «صدی البلد» گفت: کارلوس کی روش یک دیکتاتور است. او نگاه بالا از پایین در تیم ملی مصر دارد. سرمربی پرتغالی قصد دارد محمد شریف (مهاجم الاهلی) را نابود کند. خط زدن این بازیکن دیوانه کننده است. البته این تنها خبرسازی کی روش در مصر نبود. سایت «المصری الیوم» نیز با تیتیر «بحران میان کی روش با فدراسیون به خاطر کریسمس» به اختلافات این مربی با فدراسیون پرداخت. این رسانه مصری نوشت: کارلوس کی روش قرار بود اردوی تیم ملی برای آماده شدن در رقابت های مقدماتی جام جهانی 2022 از تاریخ 28 دسامبر آغاز کند ولی به دلیل تعطیلات کریسمس اردو را به تعویق انداخت که واکنش فدراسیون فوتبال را به همراه داشت. فدراسیون فوتبال مصر با این اقدام کی روش مخالفت کرد و اختلافات بین طرفین به خاطر تعطیلات کریسمس بالا گرفته است. انتهای پیام/ n\n"

ت) پرسمان دشوار چندکلمه‌ای: المصری الیوم الاهلی

می‌دانیم کلمات بالا در همین سند 6117 وجود دارد، عملکرد را مقایسه کنیم:

نتایج QueryProcessor:

```
ormation.html?eval=1&object=query&processsing=test.php
normal cosine search spent time: 0.0, result is:
DocID in json: 6117, Score: 0.976523432945281, Title: «روت‌انکید» شوریك URL: h
ttps://www.farsnews.ir/news/14001005000165/سم‌سیرک‌رطاخ‌ه‌ب‌ا‌ه‌ی‌رمم‌-‌اب‌یل‌اغ‌ت‌ر‌پ‌د‌رم‌خ‌ال‌ت‌خ‌ا‌-‌ت‌ف‌ر‌گ‌ب‌ق‌ل‌-‌ر‌و‌ت‌ا‌ن‌ک‌ی‌د‌-‌ش‌و‌ر‌ی‌ک/
****

DocID in json: 6705, Score: 0.976523432945281, Title: سیک‌+تفرگ‌ال‌اب‌رمم‌علم‌میت‌رد‌شن‌ارایتمید‌اب‌شوریک‌خ‌ال‌ت‌خ‌ا‌ URL: https://www.farsnews
.ir/news/14000927000118/
****

DocID in json: 6609, Score: 0.8149173302689336, Title: دم‌آرد‌ادم‌ه‌ب‌یدئاق‌یار‌ب‌رطخ‌گنز URL: https://www.farsnews.ir/news/140009280005
55/
****

DocID in json: 5122, Score: 0.8140420819082048, Title: سیک‌+دش‌رمم‌رد‌شوریک‌گ‌ر‌ز‌ب‌ل‌ا‌ج‌ن‌ج‌ین‌اب‌رق‌م‌ج‌ر‌تم URL: https://www.farsnews.ir/news/1
4001019000156/
****

DocID in json: 5532, Score: 0.8140420819082048, Title: سیک‌+دروخ‌یم‌ن‌ام‌ل‌ابتوف‌د‌رد‌ه‌ب‌وا‌؛شوریک‌زا‌رمم‌قباس‌ری‌زو‌د‌ا‌ق‌تن‌ا‌ URL: https://www
.farsnews.ir/news/14001013000274/
****

DocID in json: 6328, Score: 0.8140420819082048, Title: سیک‌+دش‌خی‌بوت‌رمم‌ل‌ابتوف‌ه‌روطس‌؛شوریک‌دی‌ج‌شل‌ا‌ج‌ URL: https://www.farsnews.ir/ne
ws/14001002000221/
****

DocID in json: 6790, Score: 0.8140420819082048, Title: سنوت‌ه‌ب‌تخ‌اب‌زا‌سپ‌شوریک‌زا‌رمم‌ن‌اتوسیک‌شپ‌دی‌ش‌ا‌ق‌تن‌ا‌ URL: https://www.farsnews
.ir/news/14000925000165/
****

DocID in json: 143, Score: 0.5773502691896258, Title: ی‌ه‌ل‌ل‌ا‌ن‌زع‌ه‌دن‌ی‌آ‌زا‌ت‌ب‌ج‌س‌و‌د‌الوف‌یار‌ب‌ن‌اشن‌و‌ط‌خ‌ا‌ت‌ل‌الق‌ت‌س‌ا‌رد‌ین‌و‌ج‌ام‌ارت‌س‌ا‌ه‌رطاخ‌زا‌ URL: https://www.farsnews.ir/news/14001222000300/
****

DocID in json: 155, Score: 0.5773502691896258, Title: درک‌ت‌فا‌ی‌ه‌ل‌ل‌ارون‌+ت‌ف‌گ‌ر‌ب‌ج‌و‌ا‌ی‌اه‌زور‌ه‌ب‌یدئاق‌ا‌ه‌رن‌وی‌ژل‌درک‌لم‌ع‌یب‌ای‌ز‌ا‌ URL: https://
www.farsnews.ir/news/14001222000096/
****

DocID in json: 163, Score: 0.5773502691896258, Title: م‌م‌درکن‌ل‌وب‌ی‌ق‌/م‌دش‌خ‌ز‌رب‌یدم‌ج‌مل‌گ‌یی‌اد‌ج‌اب‌:ین‌ی‌س‌ج‌ا‌ن‌اری‌ای‌ح‌یرم‌ن‌یرت‌ت‌ک‌اس‌اب‌و‌گ‌و‌ت‌ف‌گ URL: https://www.farsnews.ir/news/14001221000302/
****
```

نتایج Efficient:

```

champ list + normal cosine search spent time: 0.0009968280792236328, result is:
DocID in json: 155, Score: 1.0, Title: دەرگه‌ی‌ه‌زور ه‌ب‌یدئاق|ه‌رن‌ویژل درکلمع‌یب‌ایزرا URL: https://www.farsnews.ir/news/14001222000096/
دەرگه‌تفا-یه‌لل‌ارون-تشیگرب‌ج‌وا-یه‌زور-ه‌ب‌یدئاق|ه‌رن‌ویژل-درکلمع-یب‌ایزرا
****

DocID in json: 180, Score: 1.0, Title: یه‌لل‌ارون و‌یدئاق‌اب‌یله‌ال‌بابیش‌تشیگش|تارام‌ال‌ایتوف‌گیل URL: https://www.farsnews.ir/news/1400121001027/
یه‌لل‌ارون-و-یدئاق-اب-یله‌ال‌بابیش-تشیگش|تارام‌ال‌ایتوف-گیل
****

DocID in json: 807, Score: 1.0, Title: یه‌رن‌ویژل درکلمع‌یب‌ایزرا |نکیزاب‌نیرته‌ب‌یه‌لل‌ارون |تصرف‌یدئاق/دش‌یله‌ال‌نکیزاب‌نیرته‌ب‌یه‌لل‌ارون URL: https://www.farsnews.ir/news/14001214000126/
تصرف-یدئاق-دش-یله‌ال‌نکیزاب-نیرته‌ب-یه‌لل‌ارون-|ه‌رن‌ویژل-درکلمع-یب‌ایزرا
****

DocID in json: 812, Score: 1.0, Title: ملیف‌دشن‌ه‌دحول‌ه‌زاورد‌یه‌ار‌یه‌لل‌ارون‌شکرس‌توش/یدئاق‌یزوس‌تصرف URL: https://www.farsnews.ir/news/14001214000095/
ملیف-دشن-ه‌دحول-ه‌زاورد-یه‌ار-یه‌لل‌ارون-شکرس-توش-یدئاق-یزوس-تصرف
****

DocID in json: 942, Score: 1.0, Title: دش‌م‌انوگن‌اب‌تباقر‌ه‌نیزگ‌اقیرف‌ان‌ام‌ره‌ق‌یجرم‌رس URL: https://www.farsnews.ir/news/14001211000432/
دش‌م‌انوگن‌اب‌تباقر‌ه‌نیزگ‌اقیرف‌ان‌ام‌ره‌ق‌یجرم‌رس
****

DocID in json: 967, Score: 1.0, Title: س‌کع‌+ایس‌آ‌ل‌اس‌درم‌اب‌س‌هل‌وپ‌س‌رپ‌قب‌اس‌ن‌کیزاب‌لباقت‌یله‌ال‌بابیش‌هراتس‌یه‌لل‌ارون URL: https://www.farsnews.ir/news/14001211000089/
ایس‌آ‌ل‌اس‌درم‌اب‌س‌هل‌وپ‌س‌رپ‌قب‌اس‌ن‌کیزاب‌لباقت‌یله‌ال‌بابیش‌هراتس‌یه‌لل‌ارون
****

DocID in json: 979, Score: 1.0, Title: یله‌ال‌ربارب‌ن‌اگد‌از‌ین‌اغ‌ن‌ن‌ارای‌یواس‌ت‌|رطق‌ن‌اگراتس‌گیل URL: https://www.farsnews.ir/news/14001210000625/
یله‌ال‌ربارب‌ن‌اگد‌از‌ین‌اغ‌ن‌ن‌ارای‌یواس‌ت‌|رطق‌ن‌اگراتس‌گیل
****

DocID in json: 1009, Score: 1.0, Title: س‌کع‌+یدئاق‌ین‌یش‌ن‌تکم‌ین‌زا‌یتارام‌ه‌ناس‌ر‌ب‌لاج‌داقتن‌ا URL: https://www.farsnews.ir/news/1400121000165/
س‌کع‌یدئاق‌ین‌یش‌ن‌تکم‌ین‌زا‌یتارام‌ه‌ناس‌ر‌ب‌لاج‌داقتن‌ا
****

DocID in json: 1219, Score: 1.0, Title: ه‌ب‌ل‌ال‌قتسا‌قب‌اس‌هراتس‌ه‌یش‌ت‌دش‌یله‌ال‌یجرم‌رس‌زا‌داقتن‌ا‌شع‌اب‌یدئاق URL: https://www.farsnews.ir/news/14001208000297/
ه‌ب‌ل‌ال‌قتسا‌قب‌اس‌هراتس‌ه‌یش‌ت‌دش‌یله‌ال‌یجرم‌رس‌زا‌داقتن‌ا‌شع‌اب‌یدئاق
****

DocID in json: 1234, Score: 1.0, Title: شادن‌م‌ادن‌ایرب‌یتصرف‌یدئاق/یله‌ال‌بابیش‌رترب‌ن‌کیزاب‌ن‌یم‌ود‌یه‌لل‌ارون |ه‌رن‌ویژل درکلمع‌یب‌ایزرا URL: https://www.farsnews.ir/news/14001208000094/
یدئاق‌یله‌ال‌بابیش‌رترب‌ن‌کیزاب‌ن‌یم‌ود‌یه‌لل‌ارون-|ه‌رن‌ویژل-درکلمع-یب‌ایزرا
****

```

نتایج Positional:

positional + normal cosine search spent time: 0.0, result is:

DocID in json: 5122, Score: PHRASE-QUERY-RESULT, Title: سیک-دش-ریم رد شوریک گرزب لاجنچ ین ابرق مچرتم URL: https://www.farsnews.ir/newss/14001019000156/ سیک-دش-ریم-رد-شوریک-گرزب-ل-اچنچ-ین-ابرق-مچرتم

DocID in json: 5532, Score: PHRASE-QUERY-RESULT, Title: سیک-دروغیم-ام لابتوف درد هب وا؛ شوریک زا ریم قباس ریزو داقتن ا URL: https://www.farsnews.ir/news/14001013000274/ سیک-دروغیم-ام-ل-ابتوف-درد-هب-وا-شوریک-زا-ریم-قباس-ریزو-د-اقتن-ا

DocID in json: 6117, Score: PHRASE-QUERY-RESULT, Title: سیک-سیرک-رطاخ هب اهیرم اب یل اغترب درم فالخا/تفرگ یقل «روتانکید» شوریک URL: https://www.farsnews.ir/news/14001005000165/ سیک-سیرک-رطاخ-هب-اهیرم-اب-یل-اغترب-درم-فالخا-تفرگ-یقل-روتانکید-شوریک

DocID in json: 6328, Score: PHRASE-QUERY-RESULT, Title: سیک-دش-خیبوت-ریم لابتوف هروطس؛ شوریک دیدج شل اچ URL: https://www.farsnews.ir/news/14001002000221/ سیک-دش-خیبوت-ریم-ل-ابتوف-هروطس-شوریک-دیدج-شل-اچ

DocID in json: 6705, Score: PHRASE-QUERY-RESULT, Title: سیک-تفرگ-ال اب ریم یم میت رد شن ارایتید اب شوریک فالخا URL: https://www.farsnews.ir/news/14000927000118/ سیک-تفرگ-ال-اب-ریم-یم-میت-رد-شن-ارایتید-اب-شوریک-فالخا

DocID in json: 6790, Score: PHRASE-QUERY-RESULT, Title: سینوت-هب-تخ-اب-زا سپ-شوریک-زا ریم ناتوسکشپ دیدش داقتن ا URL: https://www.farsnews.ir/news/14000925000165/ سینوت-هب-تخ-اب-زا-سپ-شوریک-زا-ریم-ناتوسکشپ-دیدش-د-اقتن-ا

DocID in json: 6609, Score: 0.8149173302689336, Title: دم آرد ادم هب یدئاق یارب رطخ گنز URL: https://www.farsnews.ir/news/14000928000555/ دم-آرد-ادم-هب-یدئاق-یارب-رطخ-گنز

DocID in json: 143, Score: 0.5773502691896258, Title: یهلل اترع هدنیآ زا تبجس و دالوف یارب نائش و طخ ات لالقنسا رد ینوج ام ارتسا ه رطاخ زا URL: https://www.farsnews.ir/news/14001222000300/ زا-تبجس-و-دالوف-یارب-نائش-و-طخ-ات-ل-القنسا-رد-ینوج-ام-ارتسا-ه-رطاخ-زا

DocID in json: 155, Score: 0.5773502691896258, Title: درک-تفا-یهلل-ارون-تتگرب-چوا-یاهزور-هب-یدئاق|اهرنویژل-درکلمع-یبایزرا URL: https://www.farsnews.ir/news/14001222000096/ درک-تفا-یهلل-ارون-تتگرب-چوا-یاهزور-هب-یدئاق|اهرنویژل-درکلمع-یبایزرا

DocID in json: 163, Score: 0.5773502691896258, Title: م-درکن-لویق/م-دش-یخزرب-یدم-م-لگ-ی-یادج-اب-ینیسج-اناری-یبرم-نیرت-تکاس-اب-وگ-و-تفگ URL: https://www.farsnews.ir/news/14001221000302/ م-درکن-لویق-م-دش-یخزرب-یدم-م-لگ-ی-یادج-اب-ینیسج-اناری-یبرم-نیرت-تکاس-اب-وگ-و-تفگ

توضیح: همانطور که میبینیم، اولین سند بازگردانده شده توسط QueryProcessor، دقیقاً سند 6117 است! ولی Efficient چون در شاخص‌هایش اصلاً برای کلمه المصری، posting مربوط به 6117 را وارد نکرده بوده، و نتایج متفاوتی با شاخص دیگر دارد.

عملکرد Positional هم به خوبی مشخص است، و سومین سند بازگردانده شده، سند 6117 است.

پایان