

(۱) در این بخش می‌خواهیم نحوه استخراج توکن‌ها و نرمال‌سازی آنها در پروژه شما را بررسی کنیم. لطفاً برای اولین خبر از مجموعه داده پردازش شده، متن اصلی خبر، توکن‌های استخراج شده در مرحله اول، نتیجه نرمال‌سازی توکن‌ها، حذف کلمات پرتکرار و ریشه‌یابی را به شکل مناسب نشان دهید. به عنوان مثال می‌توانید ابتدا متن اصلی خبر را خروجی دهید. سپس هر یک از توکن‌های استخراج شده که به گونه‌ای از هم جدا شده باشند را نمایش دهید و سپس هر یک از توکن‌ها که در فرآیند نرمال‌سازی تغییری می‌کنند را به دو شکل قبل از تغییر و بعد از تغییر نمایش دهید. در ادامه کلماتی از متن که حذف شده اند را مشخص کنید و نهایتاً نتیجه ریشه‌یابی روی فعل‌ها را مشخص کنید.

متن اصلی خبر:

به گزارش خبرگزاری فارس، سردار آزمون ستاره تیم ملی کشورمان و عضو باشگاه زنیت به دلیل درخشش بی نظیرش در لیگ روسیه و نزدیک بودن زمان پایان قراردادش با باشگاه روسی مشتریان زیادی پیدا کرده است. سیماک سرمربی زنیت روز گذشته در مصاحبه با رسانه های روسیه اعلام کرد مسئله آینده آزمون آسان است این بازیکن قصد تمدید قراردادش را ندارد و در تابستان به باشگاه جدیدی می رود. در همین رابطه سایت «hitc» به تحلیل حرف های سرمربی زنیت پرداخت و این حرف ها را فرصتی استثنایی برای دو باشگاه اورتون و نیوکاسل دانست. این رسانه انگلیسی، نوشت: سردار آزمون ستاره ایرانی زنیت در لیگ قهرمانان اروپا در یک بازی جذاب به کاپوس هواداران چلسی تبدیل شد. این ستاره ایرانی یک گل فوق العاده در این بازی به ثمر رساند و اگر واکنش های فوق العاده کیه آ نبود می توانست چندین بار دیگر دروازه شاگردان توخل را باز کند. ستاره ایرانی از دو سال گذشته بنا به اعلام مدیر برنامه اش از باشگاه اورتون با ریاست فرهاد مشیری ایرانی پیشنهاد داشته و یکی از گزینه های تقویت خط حمله این تیم است. از طرف دیگر نیوکاسل که در آستانه سقوط قرار دارد برای رهایی از این وضعیت بد می خواهد با مدیریت جدید سعودی ها سردار آزمون را جذب کند تا خط حمله اش جانی دوباره بگیرد. قرارداد سردار آزمون تابستان پیش رو با زنیت به پایان می رسد و این بازیکن 17 میلیون پوندی به صورت رایگان به تیمی دیگر خواهد رفت. نیوکاسل و اورتون فرصتی طلایی برای جذب این بازیکن به صورت رایگان را دارند. البته مهاجم ملی پوش ایرانی از سوی باشگاه های یوونتوس، لیون و بایر لورکوزن نیز مورد توجه است و باید دید این بازیکن ایرانی 26 ساله چه تصمیمی برای آینده اش در تابستان می گیرد. آزمون از سال 2019 برای زنیت بازی می کند. ستاره ایرانی در مدت کمتر از 3 سال 3 بار قهرمانی لیگ برتر روسیه را تجربه کرد و یک بار نیز جایزه آقای گلی لیگ را از آن کرد. انتهای پیام / NBSP

توکن های استخراج شده در مرحله اول:

['به', 'گزارش', 'خبرگزاری', 'فارس', '،', '،', 'سردار', 'آزمون', 'ستاره', 'تیم', 'ملی', 'کشورمان', 'و', 'عضو', 'باشگاه', 'زنیت', 'به', 'دلیل', 'درخشش', 'بی', 'نظیرش', 'در', 'لیگ', 'روسیه', 'و', 'نزدیک', 'بودن', 'زمان', 'پایان', 'قراردادش', 'با', 'باشگاه', 'روسی', 'مشتریان', 'زیادی', 'پیدا', 'کرده‌است', '،', 'سیماک', 'سرمربی', 'زنیت', 'روز', 'گذشته', 'در', 'مصاحبه', 'با', 'رسانه', 'های', 'روسیه', 'اعلام', 'کرد', 'مسئله', 'آینده', 'آزمون', 'آسان', 'است', 'این', 'بازیکن', 'قصد', 'تمدید', 'قراردادش', 'را', 'ندارد', 'و']

'در'، 'تابستان'، 'به'، 'باشگاه'، 'جدیدی'، 'می'، 'رود'، '،'، 'در'، 'همین'، 'رابطه'، 'سایت'، '»'، 'hitc'، '«'، 'به'، 'تحلیل'، 'حرف'،
 'های'، 'سرمری'، 'زنیت'، 'پرداخت'، 'و'، 'این'، 'حرف'، 'ها'، 'را'، 'فرصتی'، 'استثنایی'، 'برای'، 'دو'، 'باشگاه'، 'اورتون'، 'و'،
 'نیوکاسل'، 'دانست'، '،'، 'این'، 'رسانه'، 'انگلیسی'، '،'، 'نوشت'، '،'، 'سردار'، 'آزمون'، 'ستاره'، 'ایرانی'، 'زنیت'، 'در'، 'لیگ'،
 'قهرمانان'، 'اروپا'، 'در'، 'یک'، 'بازی'، 'جذاب'، 'به'، 'کابوس'، 'هوادران'، 'چلسی'، 'تبدیل'، 'شد'، '،'، 'این'، 'ستاره'، 'ایرانی'، 'یک'،
 'گل'، 'فوق'، 'العاده'، 'در'، 'این'، 'بازی'، 'به'، 'ثمر'، 'رساند'، 'و'، 'اگر'، 'واکنش'، 'های'، 'فوق'، 'العاده'، 'کپه'، 'آ'، 'نبود'، 'می'،
 'توانست'، 'چندین'، 'بار'، 'دیگر'، 'دروازه'، 'شاگردان'، 'توخل'، 'را'، 'باز'، 'کند'، '،'، 'ستاره'، 'ایرانی'، 'از'، 'دو'، 'سال'، 'گذشته'،
 'بنا'، 'به'، 'اعلام'، 'مدیر'، 'برنامه'، 'اش'، 'از'، 'باشگاه'، 'اورتون'، 'با'، 'ریاست'، 'فرهاد'، 'مشیری'، 'ایرانی'، 'پیشنهاد'، 'داشته'، 'و'،
 'یکی'، 'از'، 'گزینه'، 'های'، 'تقویت'، 'خط'، 'حمله'، 'این'، 'تیم'، 'است'، '،'، 'از'، 'طرف'، 'دیگر'، 'نیوکاسل'، 'که'، 'در'، 'آستانه'،
 'سقوط'، 'قرار'، 'دارد'، 'برای'، 'رهایی'، 'از'، 'این'، 'وضعیت'، 'بد'، 'می'، 'خواهد'، 'مدیریت'، 'جدید'، 'سعودی'، 'ها'، 'سردار'،
 'آزمون'، 'را'، 'جذب'، 'کند'، 'تا'، 'خط'، 'حمله'، 'اش'، 'جانی'، 'دوباره'، 'بگیرد'، '،'، 'قرارداد'، 'سردار'، 'آزمون'، 'تابستان'، 'پیش'،
 'رو'، 'با'، 'زنیت'، 'به'، 'پایان'، 'می'، 'رسد'، 'و'، 'این'، 'بازیکن'، '۱۷'، 'میلیون'، 'پوندی'، 'به'، 'صورت'، 'رایگان'، 'به'، 'تیمی'،
 'دیگر'، 'خواهد'، 'رفت'، '،'، 'نیوکاسل'، 'و'، 'اورتون'، 'فرصتی'، 'طلایی'، 'برای'، 'جذب'، 'این'، 'بازیکن'، 'به'، 'صورت'، 'رایگان'، 'را'،
 'دارند'، '،'، 'البته'، 'مهاجم'، 'ملی'، 'پوش'، 'ایرانی'، 'از'، 'سوی'، 'باشگاه'، 'های'، 'یوونتوس'، '،'، 'لیون'، 'و'، 'بایر'، 'لورکوزن'،
 'نیز'، 'مورد'، 'توجه'، 'است'، 'و'، 'باید'، 'دید'، 'این'، 'بازیکن'، 'ایرانی'، '۲۶'، 'ساله'، 'چه'، 'تصمیمی'، 'برای'، 'آینده'، 'اش'، 'در'،
 'تابستان'، 'می'، 'گیرد'، '،'، 'آزمون'، 'از'، 'سال'، '۲۰۱۹'، 'برای'، 'زنیت'، 'بازی'، 'می'، 'کند'، '،'، 'ستاره'، 'ایرانی'، 'در'، 'مدت'،
 'کمتر'، 'از'، '۳'، 'سال'، '۳'، 'بار'، 'قهرمانی'، 'لیگ'، 'برتر'، 'روسیه'، 'را'، 'تجربه'، 'کرد'، 'و'، 'یک'، 'بار'، 'نیز'، 'جایزه'، 'آقای'،
 'گلی'، 'لیگ'، 'را'، 'از'، 'آن'، 'کرد'، '،'، 'انتهای'، 'پیام'، '/'، 'xa0']

تغییرات حاصل از نرمال سازی توکن ها:

[('xα0', ''), ('۳', '۳'), ('۳', '۳'), ('۲۰۱۹', '۲۰۱۹'), ('۲۶', '۲۶'), ('۱۷', '۱۷')]

تغیرات حاصل از ریشه‌یابی:

['گزارش', 'گزار', 'خبرگزاری', 'خبرگزار', 'تیم', 'تی', 'ملی', 'مل', 'کشورمان', 'کشور', 'زنیت', 'زنی', 'درخشش', 'درخش', 'بی', 'ب', 'نظیرش', 'نظیر', 'زمان', 'زم', 'پایان', 'پای', 'قراردادش', 'قرارداد', 'روسی', 'روس', 'مشتریان', 'مشتري', 'زیادی', 'زیاد', 'کرده_است', 'کرده_اس', 'سرمربی', 'سرمرب', 'زنیت', 'زنی', 'های', 'ها', 'اعلام', 'اعلا', 'کرد', 'کرد#کن', 'آسان', 'آس', 'است', 'اس', 'قراردادش', 'قرارداد', 'ندارد', 'داشت#دار', 'تابستان', 'تابست', 'جدیدی', 'جدید', 'می', 'هست', 'سایت', 'سای', 'های', 'ها', 'سرمربی', 'سرمرب', 'زنیت', 'زنی', 'پرداخت', 'پرداخ', 'ها', 'هست', 'فرصتی', 'فرصت', 'استثنایی', 'استثنا', 'برای', 'برا', 'دانست', 'دانس', 'انگلیسی', 'انگلیس', 'نوشت', 'نوش', 'ایرانی', 'ایران', 'زنیت', 'زنی', 'قهرمانان', 'قهرمان', 'بازی', 'باز', 'هواداران', 'هوادار', 'چلسی', 'چلس', 'شد', 'شد#شو', 'ایرانی', 'ایران', 'بازی', 'باز', 'رساند', 'رساند#رسان', 'واکنش', 'واکن', 'های', 'ها', 'نبود', 'بود#است', 'می', 'هست', 'توانست', 'توانس', 'شاگردان', 'شاگرد', 'ایرانی', 'ایران', 'اعلام', 'اعلا', 'اش', 'ا', 'ریاست', 'ریاس', 'مشیری', 'مشیر', 'ایرانی', 'ایران', 'یکی', 'یک', 'های', 'ها', 'تقویت', 'تقوی', 'تیم', 'تی', 'است', 'اس', 'دارد', 'داشت#دار', 'برای', 'برا', 'رهای', 'رها', 'وضعیت', 'وضع', 'می', 'هست', 'مدیریت', 'مدیری', 'سعودی', 'سعود', 'ها', 'هست', 'اش', 'ا', 'جانی', 'جان', 'بگیرد', 'گرفت#گیر', 'تابستان', 'تابست', 'پیش', 'پی', 'زنیت', 'زنی', 'پایان', 'پای', 'می', 'هست', 'رسد', 'رسید#رس', 'پوندی', 'پوند', 'صورت', 'صور', 'رایگان', 'رایگ', 'تیمی', 'تیم', 'خواهدرفت', 'خواهدرف', 'فرصتی', 'فرصت', 'طلایی', 'طلا', 'برای', 'برا', 'صورت', 'صور', 'رایگان', 'رایگ', 'دارند', 'داشت#دار', 'مهاجم', 'مهاج', 'ملی', 'مل', 'پوش', 'پو', 'ایرانی', 'ایران', 'سوی', 'سو', 'های', 'ها', 'است', 'اس', 'ایرانی', 'ایران', 'تصمیمی', 'تصمیم', 'برای', 'برا', 'اش', 'ا', 'تابستان', 'تابست', 'می', 'هست', 'گیرد', 'گرفت#گیر', 'برای', 'برا', 'زنیت', 'زنی', 'بازی', 'باز', 'می', 'هست', 'ایرانی', 'ایران', 'مدت', 'مد', 'کمتر', 'کم', 'قهرمانی', 'قهرمان', 'برتر', 'بر', 'کرد', 'کرد#کن', 'آقای', 'آقا', 'گلی', 'گل', 'کرد', 'کرد#کن', 'انتهای', 'انتها', 'پیام', 'پیا', 'هست']

توکن‌های حذف شده در مرحله حذف کلمات پرتکرار:

['و', 'در', 'به', 'از', 'این', 'که', 'با', 'را', 'کرد#کن', 'شد#شو', 'اس', 'داشت#دار', 'برا', 'آن', 'تی', 'کشور', 'یک', 'بود#است',
بر', 'باز', 'تا', 'باید', 'فارس', 'گزار', 'پیا', 'دو', 'خبرگزار', 'انتها', 'مل', 'بازیکن', 'سال', 'قرار']

۲) برای شاخص مکانی تولید شده موارد زیر را مشخص کنید:
اندازه کل شاخص (اندازه فایل شاخص به بایت بیان کنید)

سه کلمه با طولانی‌ترین Postings list را بیان کنید (طول لیست بر اساس تعداد اسناد حاوی کلمه مد نظر است و نه تعداد دفعات تکرار کلمه در اسناد).

سه کلمه با کوتاه‌ترین Postings list را به همراه شماره اسناد و شماره مکان‌ها برای یکی از کلمات بیان کنید.

پس از حذف ۵۰ کلمه پرتکرار سائز آن برابر می‌شود با ۱۳۱۰۸۰۸ بایت.

[۲۵۹۵, 'خبرنگار', (۱۹۵۷, 'امروز', (۱۸۴۶, 'اظهار')]

که در کنار هر کلمه تعداد داکيومنت‌های آنها نیز ذکر شده است.

کلمات با کوتاه‌ترین posting list ها به همراه تعداد داک‌هایشان:

[('justice', 1), ('وحد"', (۱, '***برخ')]

توکن: وحد

آیدی داک: ۱۲۰۵۰

مکان‌ها: [۳۳۴]

۳) در محاسبه بردار tf-idf اسناد مطلوب است:

طول بردار tf-idf هر سند (تعداد کلمات فرهنگ لغات) چند است؟

کلمه با بیشترین و کلمه با کمترین مقادیر idf به همراه مقدار idf آنها را گزارش کنید.

برای اولین سند از مجموعه اسناد، کلمه با بیشترین و کلمه با کمترین وزن tf-idf را بیان کنید.

شبیه‌ترین سند از میان کلیه اسناد موجود در مجموعه اسناد به اولین سند بر اساس شباهت کسینوسی را پیدا کرده و میزان شباهت کسینوسی دو سند و محتوای سند مورد نظر را ارائه کنید. تحلیل خود از نتیجه را بیان کنید.

تعداد کلمات فرهنگ لغت: ۳۰۲۸۷

کلمه PDI صادر بیشترین مقدار ۸.۵۱۷۱۹۳۱۹۱۴۱۶۲۳۸ را دارد.

کلمه خبرنگار کمترین مقدار ۰.۶۵۵۸۵۱۳۹۵۸۱۶۲۴۸۴ را دارد.

برای داکيومنت اول داریم:

token with most idf is کپه with the idf 8.517193191416238

token with least idf is کند with the idf 1.0578542961959418

شبیه‌ترین سند دارای لینک <https://www.farsnews.ir/news/14001101000129> می‌باشد. که شباهت

دارند. ۰.۶۶۹۱۵۱۵۶۹۹۵۷۸۲۸۵

محتوای سند:

به گزارش خبرگزاری فارس، نیکولو شیرا خبرنگار سرشناس ایتالیایی درباره ماجرای پیشنهاد باشگاه میلان به سردار آزمون مهاجم ایرانی زنیت سن پترزبورگ پرده برداشت. این خبرنگار ایتالیایی اعلام کرد مدیران باشگاه میلان برنامه ای برای جذب سردار آزمون مهاجم ایرانی زنیت روسیه در تابستان پیش رو ندارند. روسونری چندین بازیکن برای تقویت خط حمله اش در نظر دارد ولی نام ستاره ایرانی زنیت در لیست میلان وجود ندارد. قرارداد ستاره ایرانی با تیم روسی در تابستان پیش رو به پایان می رسد و می تواند به هر باشگاهی به صورت رایگان برود. آزمون در این فصل ۱۵ بازی در لیگ برتر روسیه انجام داده و ۷ گل برای زنیت به ثمر رسانده است. البته ستاره ایرانی بنا به گزارش رسانه های اروپا از نیوکاسل، لیون، یوونتوس و لورکوزن پیشنهاد دارد. آزمون هیچ کدام از پیشنهادات غیر از باشگاه لیون را تایید نکرده و اصلی ترین مشتری مهاجم ملی پوش کشورمان بنا به گفته خودش لیون است. انتهای پیام/

تحلیل:

هر دو این خبرها مربوط به امکان جابجایی سردار آزمون از تیم روسی زنیت به میلان هستند که بدلیل شباهت فراوان در موضوع خبر، شخصیت اصلی خبر و تیمهای خاص نام برده شده دارای امتیاز بسیار بالایی هستند.

۴) با یک مثال (یک کوئری دلخواه) تاثیر نرمال سازی طول بردار اسناد در نتایج بازیابی شده را بررسی کنید. یک کوئری انتخاب کنید، نتایج حاصل از بازیابی را برای دو حالتی که بردار اسناد نرمال سازی شود و یا نرمال سازی نشود مقایسه کنید و تحلیل خود از نتیجه را بیان کنید.

کوئری: تاثیر نوسانات قیمت دلار در بازار کالاهای اساسی

در صورت عدم نرمال سازی داکيومنت‌های طولانی‌تر امتیاز و در نتیجه رنک بالاتری می‌گیرند زیرا شانس وقوع ترم‌های مشترک بیشتری با کوئری را دارند. در حالیکه لزوما مرتبط‌تر نیست اند:

```
Please enter the query:
تاثیر نوسانات قیمت دلار در بازار کالاهای اساسی

1. doc length: 8950, point: 49.391682066536326, title: ... URL: https://www.farsnews.ir/...
2. doc length: 27096, point: 47.11050767999201, title: ... URL: https://www.farsnews.ir/news/1400...
3. doc length: 34545, point: 46.28866376628058, title: ... URL: https://www.farsnews.ir/news/14001011000203/...
4. doc length: 14847, point: 40.88564771656267, title: ... URL: https://www.farsnews.ir/news/1400...
5. doc length: 13931, point: 37.955965430991526, title: ... URL: https://www.farsnews.ir/news/14001221000788/...
Press 1 to continue, or press 0 to finish

6. doc length: 3136, point: 37.41268640029626, title: ... URL: https://www.farsnews.ir/news/1400121000660/...
7. doc length: 11150, point: 34.36501162611724, title: ... URL: https://www.farsnews.ir/news/14000823000232/...
8. doc length: 2689, point: 31.971003668918915, title: ... URL: https://www.farsnews.ir/news/14000823000232/...
9. doc length: 1071, point: 31.768083033814012, title: ... URL: https://www.farsnews.ir/news/14000823000232/...
10. doc length: 2790, point: 30.821768877565557, title: ... URL: https://www.farsnews.ir/news/14000823000232/...
Press 1 to continue, or press 0 to finish

11. doc length: 7406, point: 30.649701730441798, title: ... URL: https://www.farsnews.ir/news/140009080000828/...
12. doc length: 1773, point: 30.64188698517367, title: ... URL: https://www.farsnews.ir/news/140009080000828/...
13. doc length: 3641, point: 30.08453211482533, title: ... URL: https://www.farsnews.ir/news/140009080000828/...
14. doc length: 10708, point: 29.518509094693776, title: ... URL: https://www.farsnews.ir/news/140009080000828/...
15. doc length: 1419, point: 29.180876586088104, title: ... URL: https://www.farsnews.ir/news/14001030000303/...
Press 1 to continue, or press 0 to finish

16. doc length: 1170, point: 28.263134309564006, title: ... URL: https://www.farsnews.ir/news/14000924000493/...
17. doc length: 8374, point: 27.80111796346228, title: ... URL: https://www.farsnews.ir/news/14000917000012/...
18. doc length: 28670, point: 27.4192752757097, title: ... URL: https://www.farsnews.ir/news/14001010000358/...
19. doc length: 1812, point: 26.841780948550665, title: ... URL: https://www.farsnews.ir/news/14001127001140/...
20. doc length: 1478, point: 26.676094196695875, title: ... URL: https://www.farsnews.ir/news/14000823000232/...
Press 1 to continue, or press 0 to finish
```

در حالیکه در صورت نرمال سازی داکيومنت‌ها بصورت عادلانه‌تری مقایسه می‌شوند و در نتیجه داکيومنت‌های طولانی و کوچک در برابر میزان شباهت مفاهیم خود با کوئری مقایسه می‌شوند:

```
URL: https://www.farsnews.ir/news/14080924080493/ب-درمیان‌دین/2-میلیارد دلار بیش از مجوز مجلس کالاهای اساسی وارد کرده/ مجوز 2.6 میلیارد دلای برای واردات کالای اساسی
1. doc length: 1871, point: 0.8897767887032287, title:
2. doc length: 1170, point: 0.7681021590846965, title: افزایش ارائه اقلام اساسی شد، پلدا با قیمت مناسب و مصوب در میادین
3. doc length: 1419, point: 0.715000607428199, title: عضو کمیسیون بودجه مجلس: عده‌ای به دنبال بهانه برای گران‌فروشی هستند
4. doc length: 640, point: 0.6780132331680245, title: دولت لایحه اصلاح ارز 4200 تومانی را هفته جاری به مجلس می‌فرستد
5. doc length: 1773, point: 0.6770608979586512, title: حریف و رقیب 50 میلیارد دلار با ارز 4200 تومانی در 3 سال گذشته/ احتمال تخصیص سهم ارزی برای واردات کالاهای اساسی
Press 1 to continue, or press 0 to finish
6. doc length: 1845, point: 0.6558646875318593, title: تولید کننده و وارد کننده مکلف به درج قیمت تولید یا واردات روی کالا شدند
7. doc length: 1478, point: 0.6327964046289239, title: عضو کمیسیون برنامه و بودجه: تعداد کالاهای اساسی مشمول ارز 4200 تومانی کاهش می‌یابد
8. doc length: 1367, point: 0.6114976833843963, title: حمایت از تولیدکنندگان باید منجر به عرضه کالای مرغوب با قیمت مناسب شود
9. doc length: 1720, point: 0.6106881389183477, title: انقضاء عضو هیأت رئیسه مجلس از انحصار کردن توزیع کالاهای اساسی
10. doc length: 3136, point: 0.6086985821383821, title: تصمیمات مهم ستاد تنظیم بازار برای کاهش قیمت برخی اقلام و کالاها
Press 1 to continue, or press 0 to finish
11. doc length: 598, point: 0.5806545566938715, title: تصمیم گیری درباره اصلاح ساختار بازار خودرو/ تاکید بر نظارت دقیق بر روند تامین و عرضه کالاهای اساسی
12. doc length: 955, point: 0.5751024220408791, title: استقرار پرداخت ارز توجیحی یعنی واریز 360مزار میلیارد تومان به جیب 100 نفر
13. doc length: 1812, point: 0.5718359560080188, title: برای کنترل قیمت‌ها/ دستور مجوز برای گسترش بازرسی در بازار برای کنترل قیمت‌ها
14. doc length: 1305, point: 0.5704251839770135, title: رد توییت شش‌ماهه به صفر رسید/ هیچکس انتظار ندارد تورم، شش ماهه به صفر برسد
15. doc length: 1155, point: 0.5666159135563301, title: نماینده مجلس: ارز 4200 تومانی موجب ایجاد زانیت و انحصار برای عده‌ای خاص شد
Press 1 to continue, or press 0 to finish
16. doc length: 2689, point: 0.5585303054136628, title: روایت میرکاطمی از مفکرات ارز 4200/ ارز توجیحی نهاده‌های دایمی را حذف کنید
17. doc length: 1394, point: 0.552680113761076, title: مصوبه تلفیق برای بازگشت ارز توجیحی باعث ناهمخوانی در بودجه می‌شود
18. doc length: 1790, point: 0.5292825561973982, title: نظارت کدام نهاد بر بازار میوه و گوشت باید تقویت شود؟/ ارتباط فقدان سامانه توزیع با گرانی‌ها
19. doc length: 2790, point: 0.5249266622572177, title: تخلف یکی از کمرکرها در توزیع میلیاردها دلار کالا با قیمت غیرواقعی
20. doc length: 1694, point: 0.5245986896724941, title: به‌صورت پارتانه زنجیره تولید ETS احتمال حذف دلار 4200 تومانی و پرداخت مابه‌التفاوت ارز توجیحی با
Press 1 to continue, or press 0 to finish
```

با بررسی داکيومنت ها در دو حالت متوجه می‌شویم که نرمال سازی به افزایش شباهت داکيومنت‌های برگردانده شده کمک می‌کند.

کوئری: نامہ رسمی فدراسیون فوتبال بہ تیم استقلال و تراکتور

نسخه رسمی فاینال فرماد به این شکل در اختیار شما قرار می‌گیرد:

```
documents point calculated in 0.66
1. doc_id: 6464, point: 0.607571665247582, title: اعتراض رسمی تراکتور به داوود بازی تراکتور و مع رفسنجان URL: https://www.farsnews.ir/news/1400892900795/
2. doc_id: 2012, point: 0.5641087767268521, title: با اعلام رسمی کشتزار امین فوتمال آمیا/ موزونه معالنه بهارونه و مورعلی غفرانوسنی از URL: https://www.farsnews.ir/news/1400112890087/
3. doc_id: 2760, point: 0.533095735683998, title: موافقه معصیه-آ- انفصال-کامیانه به تراکتور URL: https://www.farsnews.ir/news/1400117980876/
4. doc_id: 5106, point: 0.4883376182002858, title: سانه بدون-داشن سومرسی URL: https://www.farsnews.ir/news/140013018001928/
5. doc_id: 3255, point: 0.4572394521252221, title: گلگهر به-دنمال گلرینه پرسپولیس-در ترمیز URL: https://www.farsnews.ir/news/1400111080358/
Press 1 to continue, or press 0 to finish
1
6. doc_id: 6806, point: 0.454628566796959, title: برای-معرافه-این-تیم-ندارند URL: https://www.farsnews.ir/news/14008924080679/
7. doc_id: 3613, point: 0.3964628839839895, title: جدایی-معرافه-این-تیم-ندارند URL: https://www.farsnews.ir/news/14001066808557/
8. doc_id: 2444, point: 0.3903108338480412, title: مدافع میکان به تراکتور پیوست URL: https://www.farsnews.ir/news/1400121000538/
9. doc_id: 6528, point: 0.3834587748255803, title: بدنامی تراکتور-هم-رسید URL: https://www.farsnews.ir/news/140089292000092/
10. doc_id: 6797, point: 0.3815356641895272, title: اعتراض رسمی فدراسیون وزنه برداری ایران به اشتباه برگزاشنده معایقات فدراسی جاسی URL: https://www.farsnews.ir/news/1400092500/
Press 1 to continue, or press 0 to finish
1
11. doc_id: 701, point: 0.3750788575143987, title: تاسوای یک-نیمه از تراکتور URL: https://www.farsnews.ir/news/14001215000709/
12. doc_id: 2968, point: 0.36565178736853526, title: ترمیز تراکتور-تعطیل-شد URL: https://www.farsnews.ir/news/1400115800171/
13. doc_id: 3045, point: 0.3561882855964518, title: نیاسیاب-رئیس URL: https://www.farsnews.ir/news/1400113300579/
14. doc_id: 3389, point: 0.35953335595817375, title: ترمیزات تراکتور-آغاز-شد-خاموش URL: https://www.farsnews.ir/news/14001399006430/
15. doc_id: 3839, point: 0.35419788484163, title: کزواسی میزوند URL: https://www.farsnews.ir/news/1400105890946/
Press 1 to continue, or press 0 to finish
1
16. doc_id: 6333, point: 0.351978390084071, title: انفصال و پرسپولیس یک قدم تا حذف از آمیا/ فرست 5 روزه برای تراکتور جهت ارائه مدارک URL: https://www.farsnews.ir/news/14001082020/
17. doc_id: 1498, point: 0.33975896623337914, title: ووتر تراکتور در خانه میکان-در نیمه اول URL: https://www.farsnews.ir/news/140012040001912/
18. doc_id: 2307, point: 0.33461807064873933, title: مدافع سابق-مع-رفسنجان به تراکتور پیوست URL: https://www.farsnews.ir/news/14001123000837/
19. doc_id: 4289, point: 0.33047937271170175, title: نا-خداغلی می کند-حکم URL: https://www.farsnews.ir/news/14001828808628/
20. doc_id: 485, point: 0.3298595967995936, title: نا-انفصال فرماد از رسمی-اعضا کرد-حکم URL: https://www.farsnews.ir/news/14001717000701/
Press 1 to continue, or press 0 to finish
```

```
calculating tf-idf and documents length finished in 0.5640511512756348
champions list created in 0.5170557498931885
Please enter the query:
داده رسمی فدراسیون فوتبال به تیم استقلال و تراکتور
documents point calculated in 0.0
```

بدون استفاده از لیست قهرمانان:

```
(داده‌های پردازش شده در 2019-08-08)
documents point calculated in 0.1216654816479553
1. doc_id: 6464, point: 0.961415097306315, title: اعتراض رسمی تراکتور به داوری بازی تراکتور و مص رفسنجان, URL: https://www.farsnews.ir/news/14000929000795/
2. doc_id: 2812, point: 0.6380853333560587, title: رضا اعلام رسمی کنفدراسیون فوتبال آسیا/ هرگونه فعالیت به‌اروند و مهره‌ای غیرقانونی است, URL: https://www.farsnews.ir/news/140011280000/
3. doc_id: 3255, point: 0.6073581605791348, title: گل‌گهر به دنبال گزینه پرسپولیس در تبریز, URL: https://www.farsnews.ir/news/1400111000338/
4. doc_id: 5573, point: 0.5924086702528087, title: افس تراکتور به دور دیدار با پرسپولیس, URL: https://www.farsnews.ir/news/14001012000800/
5. doc_id: 2444, point: 0.5645016302669476, title: مدافع پیکان به تراکتور پیوست, URL: https://www.farsnews.ir/news/14001121000530/
Press 1 to continue, or press 0 to finish
↓
6. doc_id: 2760, point: 0.5330953875683998, title: موافقت مص با انتقال کیانی به تراکتور, URL: https://www.farsnews.ir/news/14001117000876/
7. doc_id: 6278, point: 0.5210137217461634, title: حلقه اضطراری در فدراسیون فوتبال برای جلوگیری از حذف استقلال و پرسپولیس از لیگ قهرمانان آسیا, URL: https://www.farsnews.ir/news/14001113000466/
8. doc_id: 3053, point: 0.5099546527575707, title: رافمنشور کنجد/ پنهان‌کاری در فدراسیون فوتبال/ متن کامل نامه مصفا را منشور کنید, URL: https://www.farsnews.ir/news/14001113000466/
9. doc_id: 3613, point: 0.49811459074573746, title: اامی, درخواست تراکتور برای جذب اواندا از لیگ برتر/ 3 بازیکن پیکان جدا می‌شوند, URL: https://www.farsnews.ir/news/14001104000537/
10. doc_id: 6797, point: 0.49349584456325085, title: اعتراض رسمی فدراسیون وزنه بودای ایران به اشتباه برگزارکننده مسابقات قهرمانی جهانی, URL: https://www.farsnews.ir/news/14000929/
Press 1 to continue, or press 0 to finish
↓
11. doc_id: 5106, point: 0.4883376182002858, title: له بدون داشتن سرمربی, اعلام اسامی تیم ملی فوتبال زیر 19 ساله بدون داشتن سرمربی, URL: https://www.farsnews.ir/news/14001018001020/
12. doc_id: 2307, point: 0.4839538901232114, title: مدافع سابق مص رفسنجان به تراکتور پیوست, URL: https://www.farsnews.ir/news/14001123000837/
13. doc_id: 1648, point: 0.4789365861511696, title: امین فوتبال/ نشست نمایندگان فراکسیون ورزش با اعضای هیات رییسه فدراسیون فوتبال, URL: https://www.farsnews.ir/news/14001202000906/
14. doc_id: 5634, point: 0.474280021223634, title: عیبه - انضباطی فراهم شده شدند, استقلال و پرسپولیس به کمیته انضباطی فراخوانده شدند, URL: https://www.farsnews.ir/news/1400101000832/
15. doc_id: 3389, point: 0.4681144972928157, title: تمرینات تراکتور آغاز شد تصاویر, URL: https://www.farsnews.ir/news/14001109000630/
Press 1 to continue, or press 0 to finish
↓
16. doc_id: 768, point: 0.46136647972458333, title: س-عکس-AFC-تمجید, از استقلال: 20 بازی بدون شکست و در مسیر قهرمانی-عکس AFC تمجید, URL: https://www.farsnews.ir/news/14001214000755/
17. doc_id: 4383, point: 0.4546675076248847, title: استقلال خواستار لغو بازی با پیکان شد, استقلال خواستار لغو بازی با پیکان شد, URL: https://www.farsnews.ir/news/14001027000760/
18. doc_id: 6806, point: 0.4546285667969529, title: برای همراهی این تیم ندانند, دو بازیکن تراکتور مشکلی برای همراهی این تیم ندارند, URL: https://www.farsnews.ir/news/14000924000879/
19. doc_id: 4289, point: 0.4480369014072982, title: سال-مد اخلاقی می‌کند-عکس, گزارش جام با کدام باشگاه از فوتبال خداحافظی می‌کند-عکس, URL: https://www.farsnews.ir/news/14001028000628/
20. doc_id: 2968, point: 0.44422112056300506, title: تمرین تراکتور تعطیل شد, تمرین تراکتور تعطیل شد, URL: https://www.farsnews.ir/news/14001119000171/
```

در صورت عدم استفاده از لیست قهرمانان زمان محاسبه امتیاز داکيومنت‌ها بدلیل لزوم محاسبه و مقایسه تمامی داکيومنت‌ها زمان بیشتری می‌گیرد ولی بدلیل بررسی تمامی آنها در خروجی، داکيومنت‌ها امتیاز بالاتری دارند.

بطور کلی با صرف زمان کمی در پیش پردازش متون می‌توان زمان پاسخ‌گویی سیستم را به کوثری‌ها بسیار بالا برد هر چند امکان دارد کمی کیفیت شباهت‌ها کمتر باشد.