**باسمه تعالی**

آزمون عملی پروژه درس بازیابی اطلاعات

نام و نام خانوادگی**: فرهاد امان**

شماره دانشجویی**: ۹۹۳۱۰۰۶**

سوال اول:

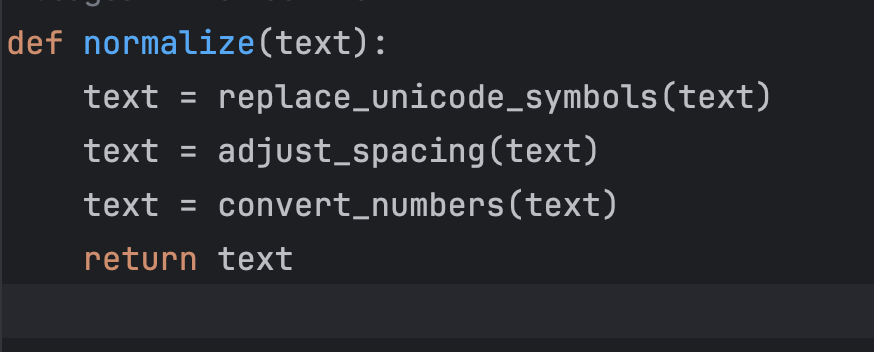
**در این سوال می‌خواهیم کلیه پیش‌پردازش‌هایی که در پروژه شما بر روی یک متن ورودی انجام می‌شود را مرور کنیم. ممکن است شما یک پیش‌پردازش را در چند مرحله انجام دهید و یا چند مرحله از پیش‌پردازش‌های خواسته شده را در یک گام انجام دهید. متناسب با نوع پیاده‌سازی خود به شکل دقیق توضیح دهید که هر یک از پیش‌پردازش‌ها را به چه صورتی پیاده‌سازی کرده اید. برای هر مورد، توضیحات خودتان را به همراه تکه کد مربوط به آن بخش و همچنین خروجی آن بخش بر روی سندی که در ادامه مشخص شده را ارائه کنید. برای پاسخگویی به این سوال از سندی که شماره آن با سه رقم سمت راست شماره دانشجویی شما یکسان است استفاده کنید. مثلا اگر شماره دانشجویی شما 40331068 است سند شماره 68 (به ترتیبی که خبرها در فایل اسناد وارد شده اند) را پردازش کنید. شماره و متن سند را در زیر وارد کنید.**

**توضیحات مربوط به پیش‌پردازش‌های انجام شده:**

**متن سند شماره ۶:**

**به گزارش خبرگزاری فارس و به نقل از  سایت باشگاه پرسپولیس، نشست خبری پیش از مسابقه سرمربیان دو تیم پرسپولیس و استقلال از هفته بیست و سوم لیگ برتر (جام خلیج فارس) با  مدیریت سازمان لیگ و هماهنگی باشگاه میزبان (پرسپولیس) در ورزشگاه شهید کاظمی به شرح زیر  برگزار می‌شود: چهارشنبه ۲۵ اسفند ساعت ۱۳ فرهاد مجیدی سرمربی استقلال ساعت ۱۳:۳۰ یحیی گل محمدی سرمربی پرسپولیس  مسابقه دو تیم روز پنجشنبه در ورزشگاه آزادی برگزار می‌شود.  به گزارش خبرگزاری فارس، پیش از این باشگاه استقلال اعلام کرده بود قرار است نشست خبری فرهاد مجیدی، سرمربی این تیم از ساعت ۱۵ در سازمان لیگ برگزار شود. انتهای پیام /**

**اولین پیش‌پردازش انجام شده نرمالایز کردن سند‌ها است**

****

**اولین اتفاقی که می‌افتد unicode\_replacements است در این بخش تعدادی از کاراکتر‌های یونیکد موجود در زبان فارسی با کلمه متناظر آن جایگزین می‌شود.**

****

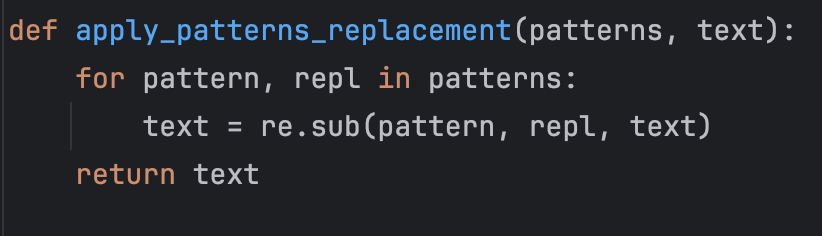
**همینطور که می‌بینید تعدادی پترن وجود دارد که در مثلا کاراکتر ریال در یونیکد با کلمه ریال جایگزین می‌شود.**

**متن سند بعد از unicode\_replacements:**

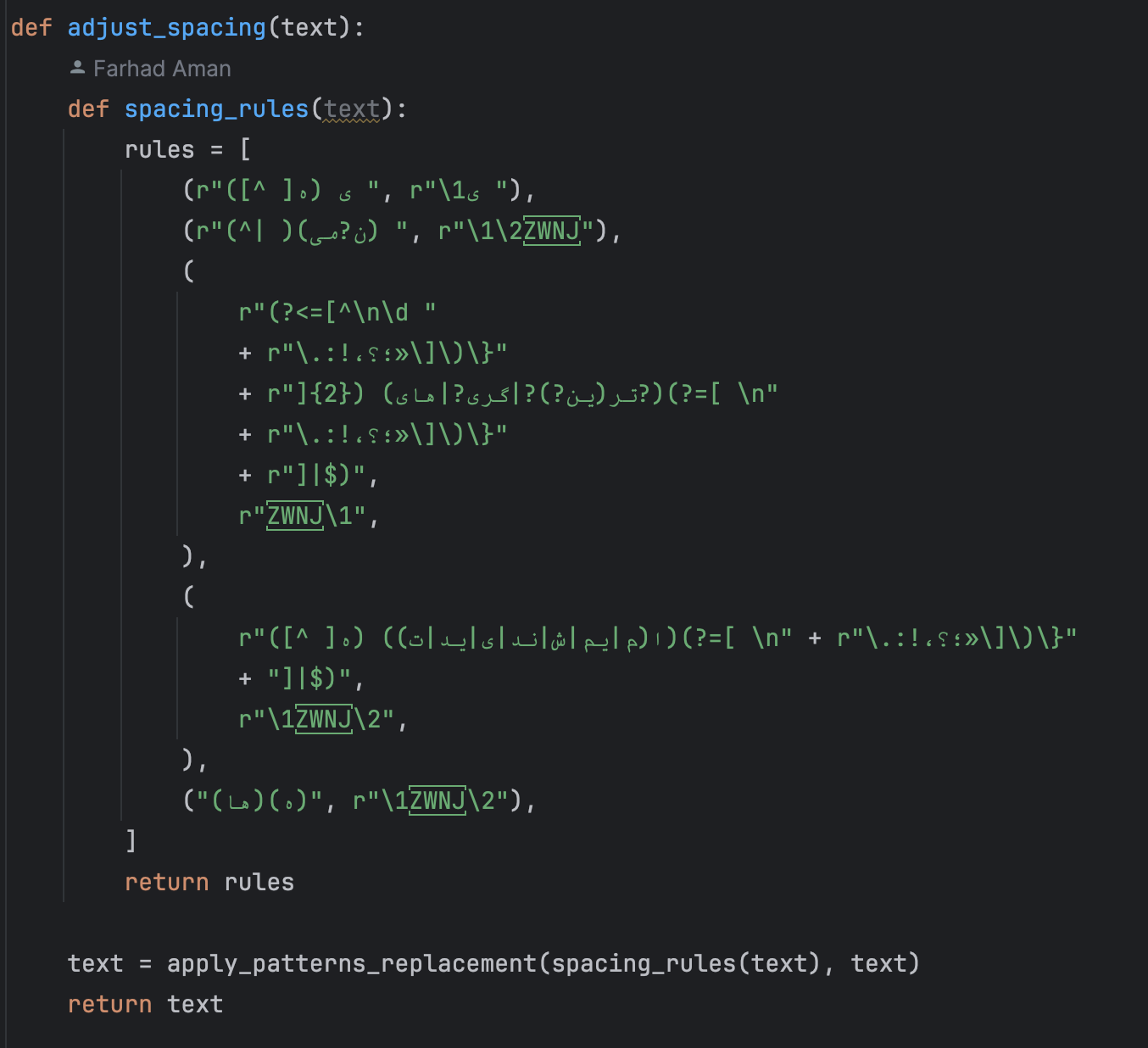
**به گزارش خبرگزاری فارس و به نقل از  سایت باشگاه پرسپولیس، نشست خبری پیش از مسابقه سرمربیان دو تیم پرسپولیس و استقلال از هفته بیست و سوم لیگ برتر (جام خلیج فارس) با  مدیریت سازمان لیگ و هماهنگی باشگاه میزبان (پرسپولیس) در ورزشگاه شهید کاظمی به شرح زیر  برگزار می‌شود: چهارشنبه ۲۵ اسفند ساعت ۱۳ فرهاد مجیدی سرمربی استقلال ساعت ۱۳:۳۰ یحیی گل محمدی سرمربی پرسپولیس  مسابقه دو تیم روز پنجشنبه در ورزشگاه ازادی برگزار می‌شود.  به گزارش خبرگزاری فارس، پیش از این باشگاه استقلال اعلام کرده بود قرار است نشست خبری فرهاد مجیدی، سرمربی این تیم از ساعت ۱۵ در سازمان لیگ برگزار شود. انتهای پیام /**

**چون این نوع کاراکتر در متن موجود نیست تغییری ایجاد نمی‌شود.**

**مرحله بعدی این است که فاصله‌گذاری‌ها مانند ی – ای – ها است**

****

**در اینجا یک پترن و یک متن گرفته می‌شود و پترن در متن اعمال می‌شود.**

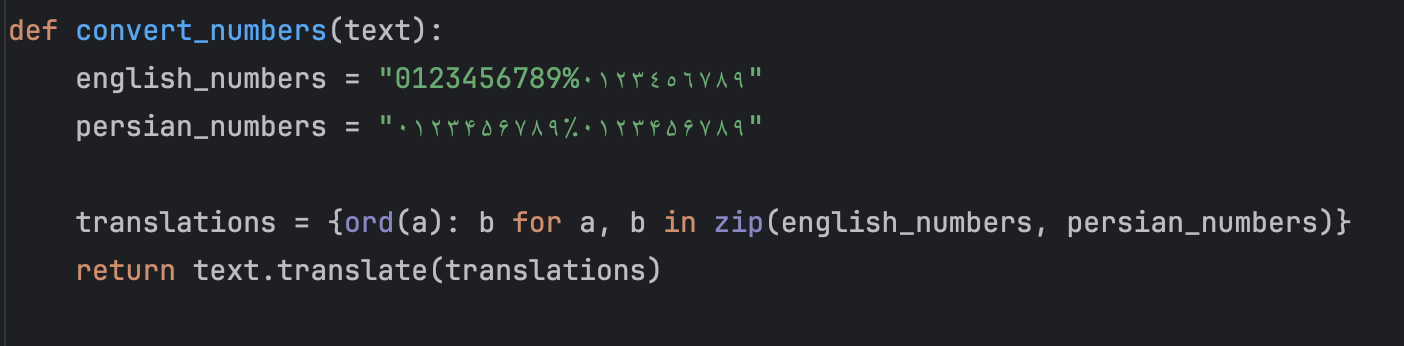
****

**در اینجا پترن‌هایی در زبان فارسی که نیاز به تغییر دارند از منابع موجود در اینترنت استخراج شده و به صورت tuple هایی ذخیره شدند و روی کد اعمال می‌شوند.**

**متن سند شماره ۶ پس از درست کردن فاصله‌گذاری‌ها:**

**به گزارش خبرگزاری فارس و به نقل از  سایت باشگاه پرسپولیس، نشست خبری پیش از مسابقه سرمربیان دو تیم پرسپولیس و استقلال از هفته بیست و سوم لیگ برتر (جام خلیج فارس) با  مدیریت سازمان لیگ و هماهنگی باشگاه میزبان (پرسپولیس) در ورزشگاه شهید کاظمی به شرح زیر  برگزار می‌شود: چهارشنبه ۲۵ اسفند ساعت ۱۳ فرهاد مجیدی سرمربی استقلال ساعت ۱۳:۳۰ یحیی گل محمدی سرمربی پرسپولیس  مسابقه دو تیم روز پنجشنبه در ورزشگاه ازادی برگزار می‌شود.  به گزارش خبرگزاری فارس، پیش از این باشگاه استقلال اعلام کرده بود قرار است نشست خبری فرهاد مجیدی، سرمربی این تیم از ساعت ۱۵ در سازمان لیگ برگزار شود. انتهای پیام /**

**مرحله بعدی تبدیل اعداد انگلیسی به فارسی است:**

****

**در اینجا تمام اعداد انگلیسی و تناظر فارسی آن‌ها موجود است و با استفاده از تابع translation تعریف شده و استفاده از متد translation روی str موجود این مرحله انجام می‌شود.**

**متن سند شماره ۶ پس از تبدیل اعداد انگلیسی به فارسی:**

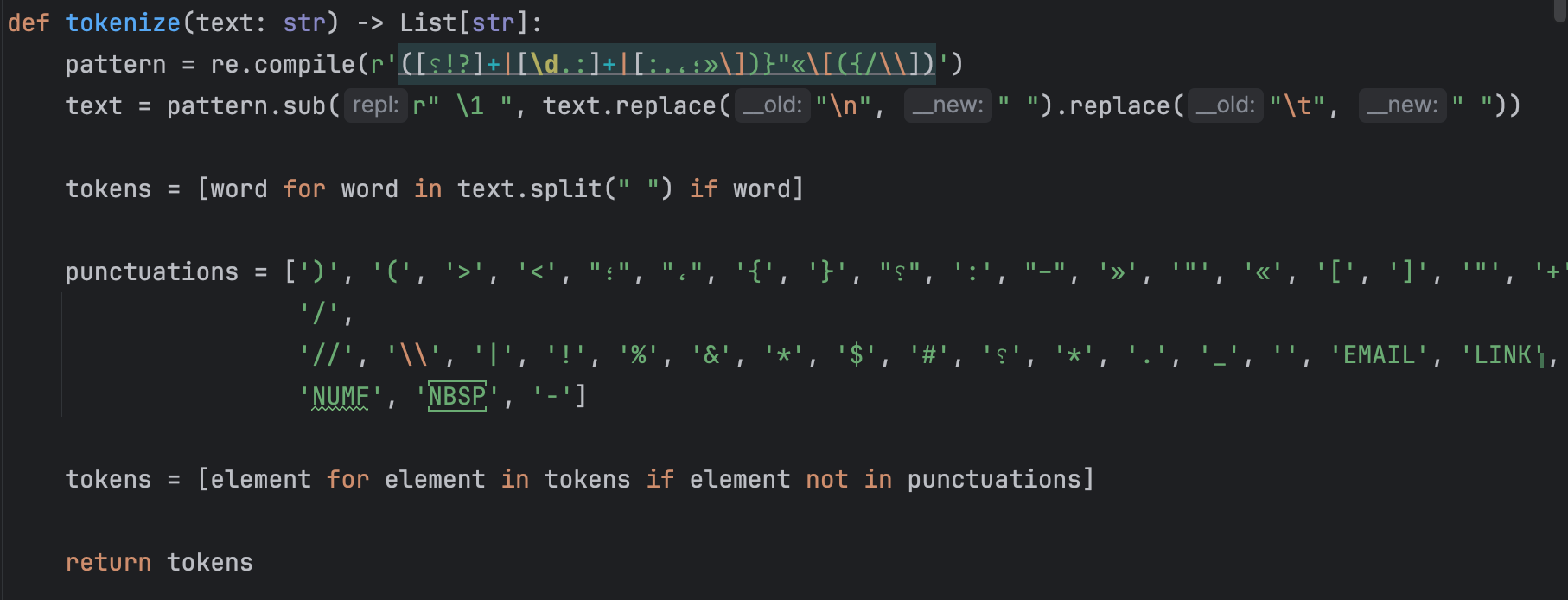
**به گزارش خبرگزاری فارس و به نقل از  سایت باشگاه پرسپولیس، نشست خبری پیش از مسابقه سرمربیان دو تیم پرسپولیس و استقلال از هفته بیست و سوم لیگ برتر (جام خلیج فارس) با  مدیریت سازمان لیگ و هماهنگی باشگاه میزبان (پرسپولیس) در ورزشگاه شهید کاظمی به شرح زیر  برگزار می‌شود: چهارشنبه ۲۵ اسفند ساعت ۱۳ فرهاد مجیدی سرمربی استقلال ساعت ۱۳:۳۰ یحیی گل محمدی سرمربی پرسپولیس  مسابقه دو تیم روز پنجشنبه در ورزشگاه ازادی برگزار می‌شود.  به گزارش خبرگزاری فارس، پیش از این باشگاه استقلال اعلام کرده بود قرار است نشست خبری فرهاد مجیدی، سرمربی این تیم از ساعت ۱۵ در سازمان لیگ برگزار شود. انتهای پیام /**

استخراج توکن**: بخش استخراج توکن مدل خود را در این بخش ارائه کنید.**

توضیحات:

**در این بخش چند تغییرات اعمال شده شامل حذف punctuation ها از توکن‌ها و همچنین حذف فاصله – نیم‌فاصله – tab – خط جدید اتفاق می‌افتد.**

کد استخراج توکن:

****

توکن‌های استخراج شده برای سند مورد نظر (هر توکن به صورت جداگانه مشخص باشد):

**['به', 'گزارش', 'خبرگزاری', 'فارس', 'و', 'به', 'نقل', 'از\xa0', 'سایت', 'باشگاه', 'پرسپولیس', 'نشست', 'خبری', 'پیش', 'از', 'مسابقه', 'سرمربیان', 'دو', 'تیم', 'پرسپولیس', 'و', 'استقلال', 'از', 'هفته', 'بیست', 'و', 'سوم', 'لیگ', 'برتر', 'جام', 'خلیج', 'فارس', 'با\xa0', 'مدیریت', 'سازمان', 'لیگ', 'و', 'هماهنگی\xa0باشگاه', 'میزبان', 'پرسپولیس', 'در', 'ورزشگاه', 'شهید', 'کاظمی', 'به', 'شرح', 'زیر\xa0', 'برگزار', 'می\u200cشود', 'چهارشنبه', '۲۵', 'اسفند', 'ساعت', '۱۳', 'فرهاد', 'مجیدی', 'سرمربی', 'استقلال', 'ساعت', '۱۳:۳۰', 'یحیی', 'گل', 'محمدی', 'سرمربی', 'پرسپولیس', '\xa0مسابقه', 'دو', 'تیم', 'روز', 'پنجشنبه', 'در', 'ورزشگاه', 'ازادی', 'برگزار', 'می\u200cشود', '\xa0به', 'گزارش', 'خبرگزاری', 'فارس', 'پیش', 'از', 'این', 'باشگاه', 'استقلال', 'اعلام', 'کرده', 'بود', 'قرار', 'است', 'نشست', 'خبری', 'فرهاد', 'مجیدی', 'سرمربی', 'این', 'تیم', 'از', 'ساعت', '۱۵', 'در', 'سازمان', 'لیگ', 'برگزار', 'شود', 'انتهای', 'پیام']**

نرمالسازی متن**:**

توضیحات:

**در این بخش ابتدا درست کردن افعال اتفاق می‌افتد:**

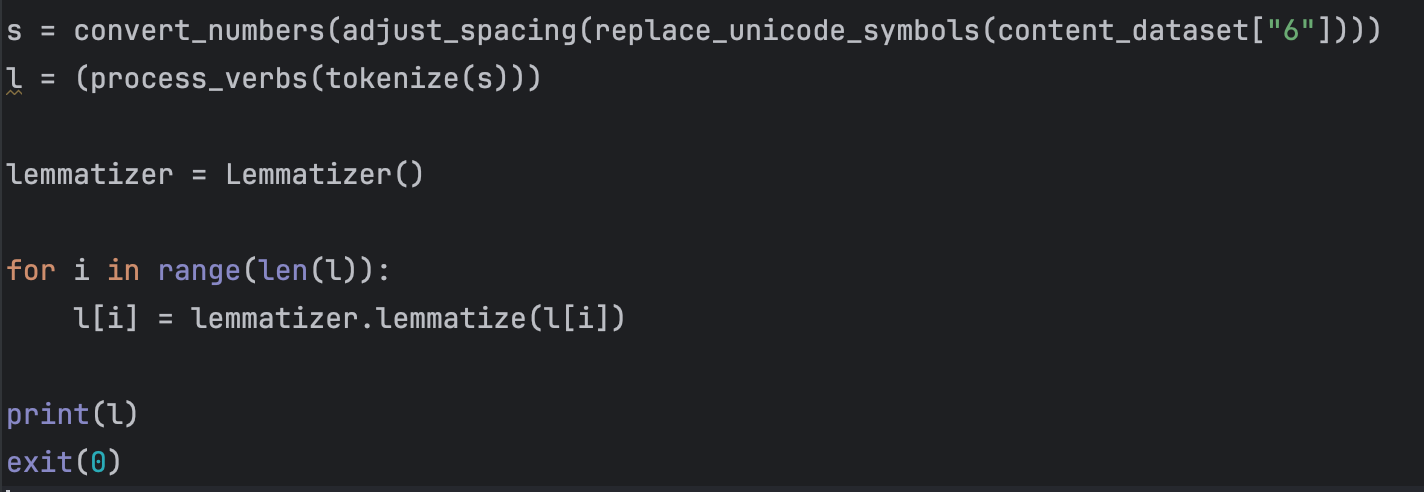
****

**در این بخش این پیشوندهای این افعال لیست شده‌اند.**

**لیست توکن‌ها بعد انجام این تغییر برای سند شماره ۶:**

**['به', 'گزارش', 'خبرگزاری', 'فارس', 'و', 'به', 'نقل', 'از\xa0', 'سایت', 'باشگاه', 'پرسپولیس', 'نشست', 'خبری', 'پیش', 'از', 'مسابقه', 'سرمربیان', 'دو', 'تیم', 'پرسپولیس', 'و', 'استقلال', 'از', 'هفته', 'بیست', 'و', 'سوم', 'لیگ', 'برتر', 'جام', 'خلیج', 'فارس', 'با\xa0', 'مدیریت', 'سازمان', 'لیگ', 'و', 'هماهنگی\xa0باشگاه', 'میزبان', 'پرسپولیس', 'در', 'ورزشگاه', 'شهید', 'کاظمی', 'به', 'شرح', 'زیر\xa0', 'برگزار', 'می\u200cشود', 'چهارشنبه', '۲۵', 'اسفند', 'ساعت', '۱۳', 'فرهاد', 'مجیدی', 'سرمربی', 'استقلال', 'ساعت', '۱۳:۳۰', 'یحیی', 'گل', 'محمدی', 'سرمربی', 'پرسپولیس', '\xa0مسابقه', 'دو', 'تیم', 'روز', 'پنجشنبه', 'در', 'ورزشگاه', 'ازادی', 'برگزار', 'می\u200cشود', '\xa0به', 'گزارش', 'خبرگزاری', 'فارس', 'پیش', 'از', 'این', 'باشگاه', 'استقلال', 'اعلام', 'کرده', 'بود', 'قرار', 'است', 'نشست', 'خبری', 'فرهاد', 'مجیدی', 'سرمربی', 'این', 'تیم', 'از', 'ساعت', '۱۵', 'در', 'سازمان', 'لیگ', 'برگزار', 'شود', 'انتهای', 'پیام']**

**بخش بعدی lemmatize کردن توکن ها است برای این بخش از Lemmatizer موجود در کتابخانه هضم استفاده شده است. Lemmatizer در واقع همان کار stemming را انجام می‌دهد اما قوی‌تر از stemming و موارد دقیقتر و بیشتری را شامل می‌شود.**

****

**لیست توکن‌ها بعد از انجام این بخش:**

**['به', 'گزارش', 'خبرگزاری', 'فارس', 'و', 'به', 'نقل', 'از\xa0', 'سایت', 'باشگاه', 'پرسپولیس', 'نشست', 'خبری', 'پیش', 'از', 'مسابقه', 'سرمربیان', 'دو', 'تیم', 'پرسپولیس', 'و', 'استقلال', 'از', 'هفته', 'بیست', 'و', 'سوم', 'لیگ', 'برتر', 'جام', 'خلیج', 'فارس', 'با\xa0', 'مدیریت', 'سازمان', 'لیگ', 'و', 'هماهنگی\xa0باشگاه', 'میزبان', 'پرسپولیس', 'در', 'ورزشگاه', 'شهید', 'کاظمی', 'به', 'شرح', 'زیر\xa0', 'برگزار', 'شد#شو', 'چهارشنبه', '۲۵', 'اسفند', 'ساعت', '۱۳', 'فرهاد', 'مجیدی', 'سرمربی', 'استقلال', 'ساعت', '۱۳:۳۰', 'یحیی', 'گل', 'محمدی', 'سرمربی', 'پرسپولیس', '\xa0مسابقه', 'دو', 'تیم', 'روز', 'پنجشنبه', 'در', 'ورزشگاه', 'ازاد', 'برگزار', 'شد#شو', '\xa0به', 'گزارش', 'خبرگزاری', 'فارس', 'پیش', 'از', 'این', 'باشگاه', 'استقلال', 'اعلام', 'کرده', 'بود#است', 'قرار', 'است', 'نشست', 'خبری', 'فرهاد', 'مجیدی', 'سرمربی', 'این', 'تیم', 'از', 'ساعت', '۱۵', 'در', 'سازمان', 'لیگ', 'برگزار', 'شد#شو', 'انت', 'پیام']**

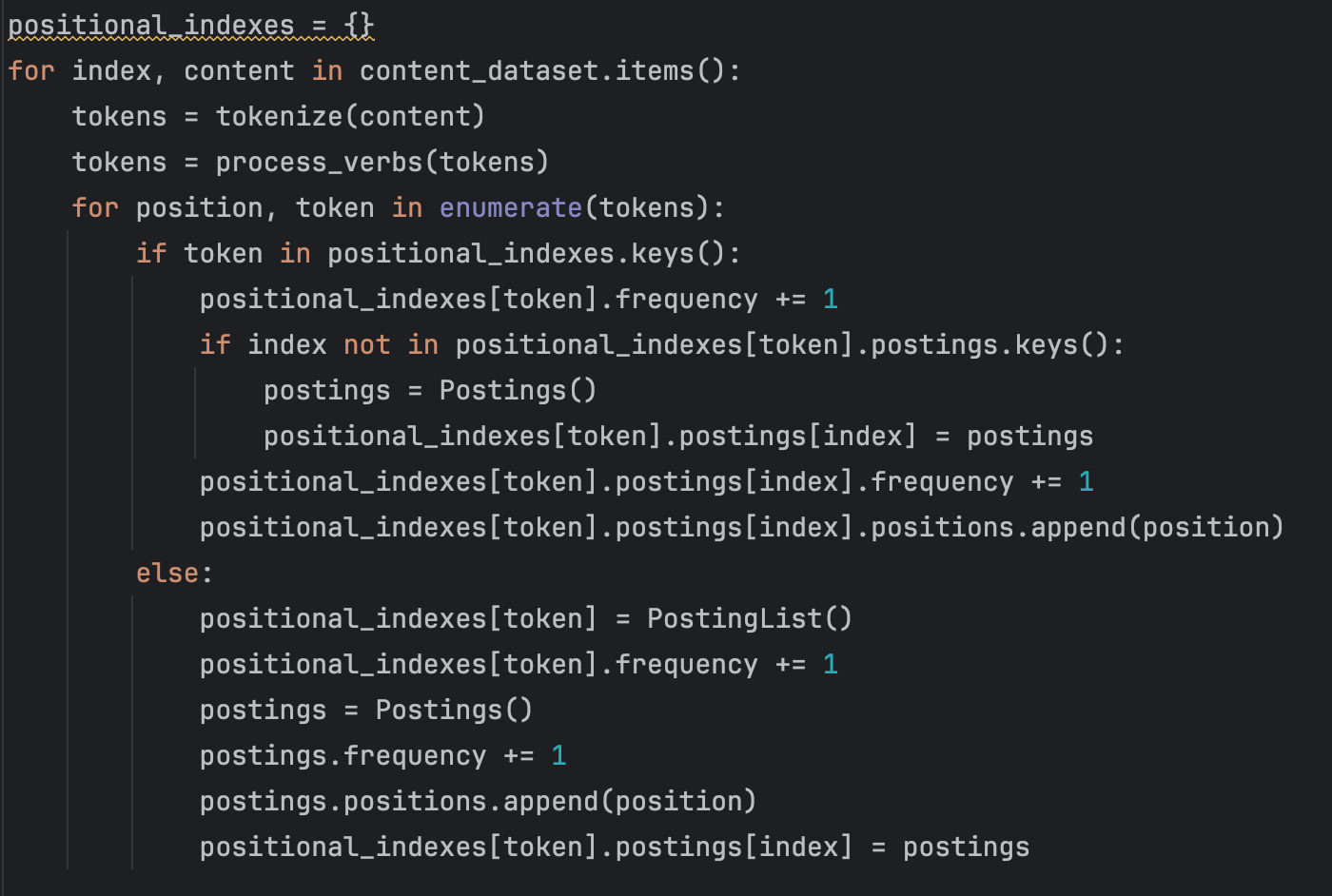
**لیست کلمات تغییر یافته پس از انجام lemmatizing:**

**[['می\u200cشود', 'شد#شو'], ['ازادی', 'ازاد'], ['می\u200cشود', 'شد#شو'], ['بود', 'بود#است'], ['شود', 'شد#شو'], ['انتهای', 'انت']]**

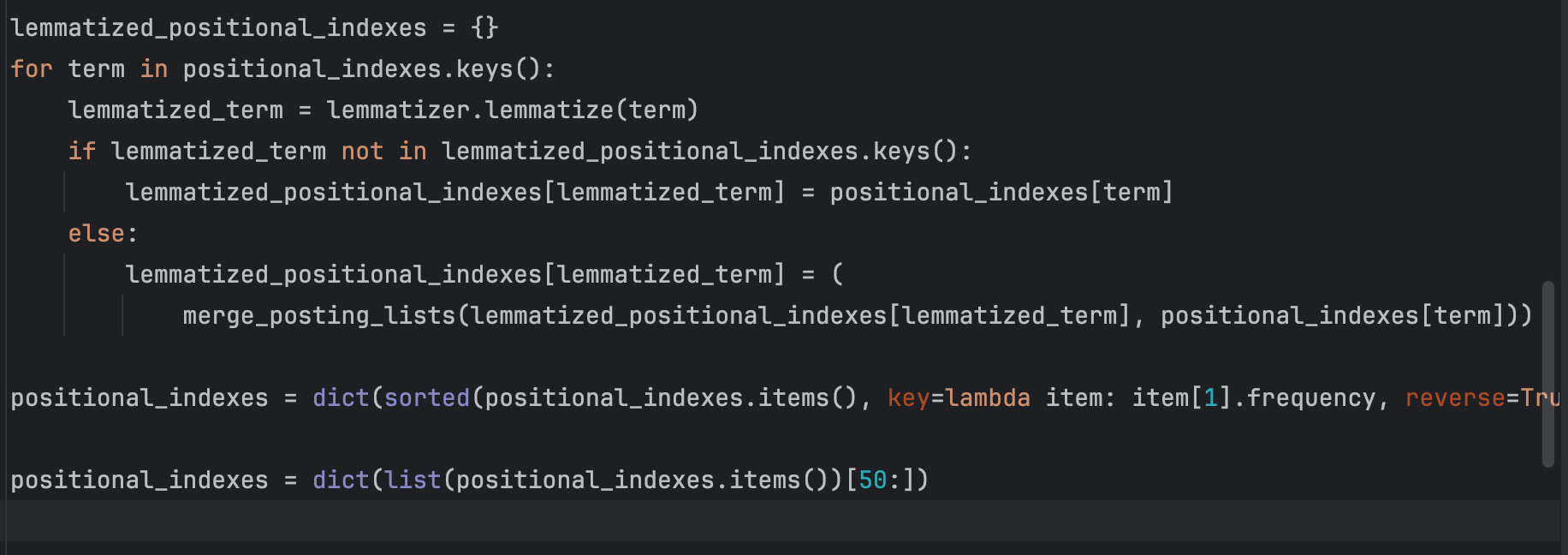
حذف کلمات پر تکرار**:**

توضیحات:

**در این بخش ابتدا positional index ساخته می‌شود.**

****

**سپس positional inex های lemmatized شده ایجاد شده و ۵۰ کلمه پرکاربرد حذف می‌شود.**

****

**لیست ۵۰ کلمه پرکاربرد:**

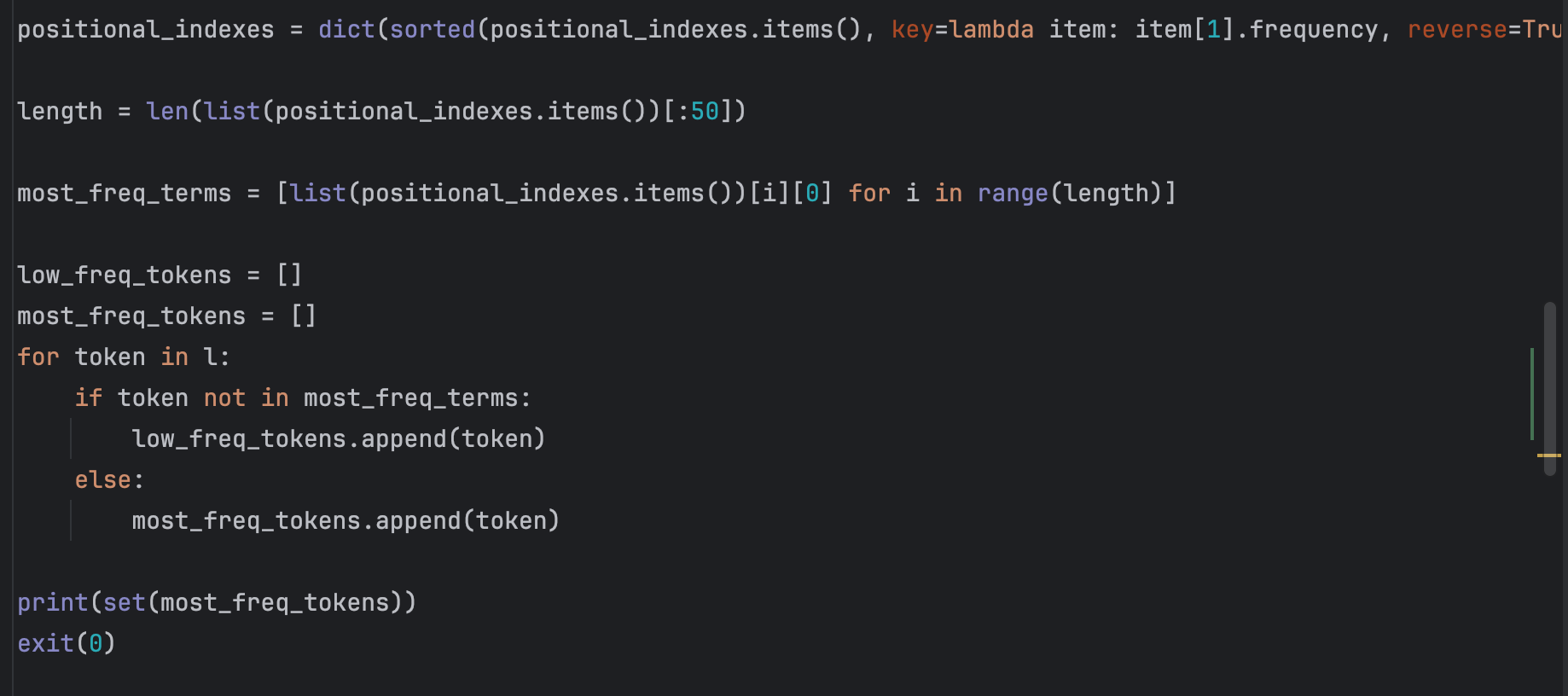
**['و', 'در', 'به', 'از', 'این', 'که', 'با', 'را', 'است', 'برای', 'کرد', 'هم', 'تیم', 'ما', 'یک', 'ان', 'شد', 'باید', 'تا', 'کشور', 'بر', 'وی', 'بازی', 'بود', 'خود', 'شده', 'مجلس', 'اسلامی', 'فارس', 'گفت', 'مردم', 'گزارش', 'پیام', 'ایران', 'خبرگزاری', 'انتهای', 'اما', 'دولت', 'شود', 'دارد', 'سال', 'ملی', 'اینکه', 'داشت', 'قرار', 'دو', 'رئیس', 'کند', 'می\u200cشود', 'کار']**

لیست کلمات سند شماره ۶ بعد از حذف ۵۰ کلمه پرکاربرد:

**['نقل', 'از\xa0', 'سایت', 'باشگاه', 'پرسپولیس', 'نشست', 'خبری', 'پیش', 'مسابقه', 'سرمربیان', 'پرسپولیس', 'استقلال', 'هفته', 'بیست', 'سوم', 'لیگ', 'برتر', 'جام', 'خلیج', 'با\xa0', 'مدیریت', 'سازمان', 'لیگ', 'هماهنگی\xa0باشگاه', 'میزبان', 'پرسپولیس', 'ورزشگاه', 'شهید', 'کاظمی', 'شرح', 'زیر\xa0', 'برگزار', 'شد#شو', 'چهارشنبه', '۲۵', 'اسفند', 'ساعت', '۱۳', 'فرهاد', 'مجیدی', 'سرمربی', 'استقلال', 'ساعت', '۱۳:۳۰', 'یحیی', 'گل', 'محمدی', 'سرمربی', 'پرسپولیس', '\xa0مسابقه', 'روز', 'پنجشنبه', 'ورزشگاه', 'ازاد', 'برگزار', 'شد#شو', '\xa0به', 'پیش', 'باشگاه', 'استقلال', 'اعلام', 'کرده', 'بود#است', 'نشست', 'خبری', 'فرهاد', 'مجیدی', 'سرمربی', 'ساعت', '۱۵', 'سازمان', 'لیگ', 'برگزار', 'شد#شو', 'انت']**

**لیست کلمات پرکاربرد حذف شده از سند:**

**{'به', 'از', 'و', 'خبرگزاری', 'تیم', 'گزارش', 'فارس', 'قرار', 'این', 'دو', 'در', 'پیام', 'است'}**

****

**در کد همانطور که مشخص است ابتدا positional index ها برحسب freq نزولی مرتب شده و سپس لیستی از ۵۰ تای اول آن گرفته و از سند شماره ۶ حذف می‌کنیم.**

سوال دوم:

**متن زیر را به عنوان یک خبر (سند) جدید به مجموعه اسناد پروژه اضافه کنید و بعد شاخص مکانی را برای همه اسناد تولید کنید و بازنمائی برداری این سند (یعنی وزن tf-idf کلیه کلمات موجود در متن داده شده) را گزارش نمایید. طبیعی است که وزن کلمات برای نسخه پیش‌پردازش شده سند محاسبه می شود که ممکن است بعضی کلمات در آن حذف شده باشند. برای نمایش بردار بازنمائی سند، هر کلمه از کلمات سند را به همراه وزن محاسبه شده برای آن کلمه گزارش کنید.**

**همچنین کلمات با بیشترین و کمترین وزن را مشخص و در مورد علت کم و زیاد بودن وزن آنها توضیح دهید.**

**دانشجویان دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر در مسابقات بین المللی برنامه سازی دانشجویی منطقه ای غرب آسیا به مقام سوم دست یافتند. به گزارش روابط عمومی دانشگاه صنعتی امیرکبیر: در بیست و چهارمین دوره از مسابقات بین‌المللی برنامه‌نویسی دانشجویی ICPC در منطقه‌ی غرب آسیا که با حضور ۸۰ تیم از دانشگاه‌های مختلف کشور در دانشگاه صنعتی شریف برگزار شد؛ تیم برنامه نویسی دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر موفق به کسب مقام سوم دانشگاهی و مقام دوم تیمی و مدال نقره در این دوره از مسابقات شد.**

**لیست کلمات موجود در سند به همراه وزن محاسبه شده برای آنها:**

**term\_id: حضور, tf\_idf: 0.03597277055763795**

**term\_id: دست, tf\_idf: 0.0494218607599093**

**term\_id: برگزار, tf\_idf: 0.04872036640431402**

**term\_id: بین, tf\_idf: 0.058600455568285637**

**term\_id: برنامه, tf\_idf: 0.08176106808546184**

**term\_id: دوم, tf\_idf: 0.05820747240668824**

**term\_id: مسابقات, tf\_idf: 0.10046942594583202**

**term\_id: دانشگاه, tf\_idf: 0.154386398088439**

**term\_id: مختلف, tf\_idf: 0.06393819908927889**

**term\_id: کسب, tf\_idf: 0.07090460509515573**

**term\_id: اسیا, tf\_idf: 0.10700696249357951**

**term\_id: عمومی, tf\_idf: 0.07817560434059959**

**term\_id: دانشجویی, tf\_idf: 0.1614469317481314**

**term\_id: مقام, tf\_idf: 0.12817812066675507**

**term\_id: سازی, tf\_idf: 0.09005349430234255**

**term\_id: دانشجویان, tf\_idf: 0.11167919910643781**

**term\_id: موفق, tf\_idf: 0.08224788271616171**

**term\_id: سوم, tf\_idf: 0.10810587396878667**

**term\_id: دوره, tf\_idf: 0.1168471566404381**

**term\_id: روابط, tf\_idf: 0.09848886922133672**

**term\_id: تیمی, tf\_idf: 0.08973091749049135**

**term\_id: غرب, tf\_idf: 0.14275632147520767**

**term\_id: بیست, tf\_idf: 0.09710972389104423**

**term\_id: بین‌المللی, tf\_idf: 0.10739995376162631**

**term\_id: صنعتی, tf\_idf: 0.1875033949088018**

**term\_id: مدال, tf\_idf: 0.12593633237844237**

**term\_id: دانشگاهی, tf\_idf: 0.10613085330168869**

**term\_id: المللی, tf\_idf: 0.11091425124653725**

**term\_id: منطقه‌ای, tf\_idf: 0.12027939022294584**

**term\_id: شریف, tf\_idf: 0.12814109371705404**

**term\_id: ۸۰, tf\_idf: 0.12665521519728695**

**term\_id: دانشگاه‌های, tf\_idf: 0.13843957659989112**

**term\_id: مهندسی, tf\_idf: 0.1882227719914342**

**term\_id: دانشکده, tf\_idf: 0.21918520607221506**

**term\_id: نقره, tf\_idf: 0.16602357192387585**

**term\_id: چهارمین, tf\_idf: 0.15650575042188605**

**term\_id: یافتند, tf\_idf: 0.16722466267018768**

**term\_id: نویسی, tf\_idf: 0.1938257250516763**

**term\_id: امیرکبیر, tf\_idf: 0.3018613451246461**

**term\_id: منطقه‌ی, tf\_idf: 0.2375590557677911**

**term\_id: کامپیوتر, tf\_idf: 0.3245829727403494**

**term\_id: برنامه‌نویسی, tf\_idf: 0.26818768864109765**

**term\_id: ICPC, tf\_idf: 0.31452702130490084**

**کلمه کامپیوتر بیشترین وزن و کلمه حضور کمترین وزن را دارد.**

**کلمه کامپیوتر دو بار در متن آمده پس tf به نسبت بالایی دارد (به نسبت دیگر کلمات) از طرفی این کلمه در کل اسناد به عمدتا ورزشی هستند در تعداد کمی سند حضور دارد پس idf بالایی دارد.**

**کلمه حضور در کل کلمه پرکاربردی است پس idf آن به نسبت کم است. از طرفی تنها یکبار در متن حاضر شده پس tf به نسبت کمی دارد.**

سوال سوم:

**در شاخص مکانی ایجاد شده برای کل اسناد پروژه، لیست پست‌های مربوط به کلمه "دانشگاه" را بررسی و به سوالات زیر پاسخ دهید.**

**الف) شماره و عنوان خبرهایی که کلمه مذکور در آنها بیشترین و کمترین وزن را دارد را گزارش کنید.**

**بیشترین وزن:**

**doc\_id: 9669**

**آیت‌الله رشاد: هیچ زمانی به اندازه امروز قشر متدین در دانشگاه نداشته‌ایم**

**tf\_idf: 0.19328039165258876**

**کمترین وزن:**

**doc\_id: 9266**

**بُرشی از کتاب « فتنه تغلب»/ روایت‌هایی از متن و فرامتن آشوب‌های**

**tf\_idf: 0.009123078690601258**

**ب) موقعیت مکانی کلمه مذکور برای ده سند اول موجود در لیست پست‌های این کلمه را گزارش کنید(شماره و عنوان سند و مکان‌های تکرار این کلمه را ذکر کنید). در صورت تکرار کلمه در یک سند، کلیه موقعیت های کلمه در آن سند ذکر شود. درستی موقعیت‌های محاسبه شده برای یک سند از سندهای مذکور را با گزارش متن سند بررسی کنید.**

**doc\_id: 105 doc\_title: رفتار متناقض در قبال «اوکراین» و 74 سال اشغالگری صهیونیست‌ها/قهرمان اسکواش جهان صدای فلسطین شد positions: [327, 341, 357]**

**doc\_id: 154 doc\_title: ملایی: ورزش همگانی عقب‌ماندگی زیادی دارد/مشکلات اقتصادی مردم را از ورزش دور کرده positions: [1279, 1424, 1645]**

**doc\_id: 452 doc\_title: قدیمی: گدا نیستیم که 100 تومان بودجه را افزایش می‌دهند/صالحی امیری هزینه کرد تا انتخابات کمیته عقب بیفتد! positions: [392]**

**doc\_id: 974 doc\_title: روایتی جالب از تیم قهرمان لیگ برتر واترپلو/ به نام شهیدی که تیم ملی و مدال را رها کرد و رفت! positions: [36, 80, 142]**

**doc\_id: 1073 doc\_title: دانشگاه آزاد قهرمان لیگ روئینگ مردان شد positions: [46, 102]**

**doc\_id: 1307 doc\_title: زمان رقابت‌های کشتی آزاد انتخابی جام تختی مشخص شد positions: [56]**

**doc\_id: 1403 doc\_title: دبیر: شهوت قدرت داشتم برای ریاست کمیته اقدام می‌کردم/ برخی دوستان از هیات اجرایی کنار بروند positions: [82]**

**doc\_id: 1422 doc\_title: اتفاقی که مربی کشتی تونس را غافلگیر کرد/العاشوری: ایران برای ما دانشگاه است positions: [289]**

**doc\_id: 1435 doc\_title: نوازی: اگر مشکل درست نکنند، استقلال می‌تواند قهرمان شود/دربی آخرسال بازی سرنوشت سازی است positions: [468]**

**doc\_id: 1603 doc\_title: نام کاروان ایران در بازی‌های پارالمپیک زمستانی پکن ۲۰۲۲ مشخص شد positions: [235]**

**برای سند شماره ۱۰۵:**

**به گزارش خبرگزاری فارس، شعار «جدایی ورزش از سیاست» که نهادهای بین المللی ورزش آن را یکی از قوانین اصلی خود می دانند، این روزها با عملیات نظامی روسیه در اوکراین زیرپا گذاشته می شود و ثابت می کند نظام سرمایه داری هر جایی که اراده کند، نه تنها قوانین را زیر سوال می برد بلکه با چرخش 180 درجه ای، علیه آن اقدام می کند. در 20 روز گذشته نهادهای بین المللی ورزش مثل فیفا، IOC، یوفا و ... علیه روسیه موضع گیری های سختی انجام داده و بر خلاف شعار «جدایی ورزش از سیاست» عمل کردند. وقتی این شعارها و رفتارهای متناقض بیشتر بروز می کند که عملیات نظامی روسیه در اوکراین را به تجاوزگری رژیم اشغالگر به خاک فلسطین تشبیه کنیم. رژیم صهیونیستی در 74 سال گذشته بارها علیه مردم فلسطین جنگ به راه انداخته و کودکان بی دفاع را به قتل رسانده ولی همین نهادهای بین المللی ورزش که این روزها برای اوکراین سینه چاک می کنند، در بیشتر از 70 سال کاری نکردند. سکوت عجیب رسانه های غربی و نظام سرمایه داری در زمینه تجاوز رژیم صهیونیستی به خاک فلسطین و مردم بی دفاعش می تواند برای تاریخ درس بزرگی باشد. در 20 روز گذشته تمام رسانه های دنیا حامی اوکراین شدند اما یادشان رفته که رژیم صهیونیستی 74 سال است که به خاک فلسطین تجاوز کرده و هر روز علیه مردم بی دفاع آن اقدام تروریستی انجام می دهد. در این سکوت خبری قهرمانی پیدا شده تا از قلب لندن و اروپا علیه نظام سرمایه داری و تبلیغات رسانه ای برای اوکراین، در حمایت از مردم مظلوم فلسطین قدعلم کند. علی عمر فرج نام قهرمان اسکواش مصری است که برای حمایت از فلسطین بسیاری از جرایم و محرومیت ها را به جان خرید و سکوتش را شکست. بازیکن تیم ملی مصر متولد قاهره است. فرج یکی از اعجوبه های اسکواش جهان به شمار می آید و در رده بندی جهانی در رده دوم ایستاده. همسر این قهرمان نیز در این ورزش فعالیت می کند. علی فرج نخبه نیز به شمار می آید. این قهرمان حامی فلسطین، در سال 2014 از دانشگاه هاروارد در رشته مهندسی مکانیک فارغ تحصیل شد. او بهترین بازیکن اسکواش این دانشگاه به شمار می آید و تنها 2 بار شکست خورده است. این بازیکن مصری عضو تیم دانشگاه هاروارد هم به شمار می آید. ماجرا از وقتی شروع شد که دو روز پیش صدای مردم مظلوم فلسطین در بهبه جنگ روسیه و اوکراین شد. ورزشکار مصری دو روز پیش قهرمان جایزه بزرگ جهانی به میزبانی لندن شد. قهرمان مصری به جای قرار گرفتن در مسیر تبلیغات رسانه ای جنگ روسیه و اوکراین، از قلب اروپا خواهان عدالتی شد و صدای مردم ستم دیده فلسطین مقابل اشغالگران شد. فرج بعد از قهرمانی مقابل دیدگان جهان، گفت: خیلی خوب می دانم آنچه که می خواهم مطرح کنم برایم مشکلاتی را بوجود می آورد. به صورت غافلگیرکننده ممکن است این حرف هایم را دخالت دادن ورزش در سیاست برداشت کنند. فیلم اظهارات آزادخواهانه قهرمان مصری در حمایت از فلسطین در زیر می توانید مشاهده کنید. وی افزود: همه در شرایط کنونی با ملت اوکراین همدردی می کنند. جنگ در هر جای دنیا محکوم است و باعث آسیب ملت ها می شود اما ما باید به فلسطین هم توجه کنیم. آنها در دهه های گذشته از جنگ علیه شان رنج می برند. فلسطینی ها از 74 سال گذشته از سوی رژیم صهیونیستی مورد ظلم قرار گرفتند ولی هیچ کس درباره فلسطین صحبت نمی کند! من درباره فلسطینی صحبت می کنم که در 74 سال گذشته اتفاقی برایش افتاده که در اوکراین الان در حال رخ دادن است. قهرمان و ورزشکار مصری از قلب اروپا صدای مردم فلسطین شد. همچنان که این قهرمان اسکواش اشاره کرد نهادهای بین المللی ورزش این اظهاراتش را سیاسی خواهند دانست و برای او جرایمی درنظر خواهند کرد ولی در مسئله روسیه با اوکراین اقدامات این نهاد در ورزش سیاسی نمی شود! تضاد بزرگ جامعه جهانی در قبال جنگ روسیه و اوکراین در مقایسه با مسئله فلسطین بار دیگر بر پوچ بودن شعار «جدایی ورزش از سیاست» تاکید می کند. انتهای پیام/**

**ج) عناصر لیست قهرمانان محاسبه شده برای این کلمه را گزارش کنید (شماره سند و وزن کلمه در این سندها را برای 20 سند اول لیست قهرمانان ذکر کنید).**

**doc\_id: 9669 tf\_idf: 0.19328039165258876**

**doc\_id: 11063 tf\_idf: 0.1803179195182215**

**doc\_id: 10261 tf\_idf: 0.17755959684392053**

**doc\_id: 10210 tf\_idf: 0.17275979241553804**

**doc\_id: 8993 tf\_idf: 0.16975203449383539**

**doc\_id: 9614 tf\_idf: 0.1692460369948589**

**doc\_id: 1810 tf\_idf: 0.16242800306897354**

**doc\_id: 10031 tf\_idf: 0.15858122798277324**

**doc\_id: 10171 tf\_idf: 0.1477821457176813**

**doc\_id: 7385 tf\_idf: 0.1460810456425358**

**doc\_id: 9238 tf\_idf: 0.13486236212400454**

**doc\_id: 10944 tf\_idf: 0.13416575581813822**

**doc\_id: 10009 tf\_idf: 0.13390461822576782**

**doc\_id: 11866 tf\_idf: 0.12837601497342743**

**doc\_id: 11940 tf\_idf: 0.1258118428250946**

**doc\_id: 8579 tf\_idf: 0.12073943939882954**

**doc\_id: 10069 tf\_idf: 0.11857836971826916**

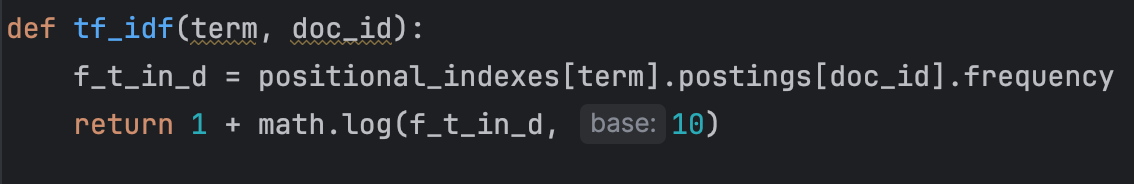
**doc\_id: 10188 tf\_idf: 0.11852789921088959**

**doc\_id: 11903 tf\_idf: 0.11807774932428054**

**doc\_id: 7497 tf\_idf: 0.1174385923092046**

سوال چهارم:

**در این سوال میخواهیم تاثیر idf در نتایج بازیابی شده را بررسی کنیم. برای این منظور شاخص جدیدی ایجاد کنید که در آن وزن کلمات در اسناد به جای رابطه tf-idf داده شده در تعریف پروژه صرفا بر اساس بخش tf محاسبه شود. با استفاده از چند عبارت جستجو و مقایسه نتایج حاصل از آنها در مدل بازیابی جدید و مدل بازیابی قبلی تاثیر این روش محاسبه بردار بازنمائی اسناد را بررسی کنید. کوئری های پیشنهادی خودتان و نتایج بازیابی شده در هر یک از دو مدل گزارش کنید و تحلیل خود را بیان نمایید. دقت داشته باشید که انتخاب عبارت جستجوی مناسب در این سوال مهم است.**

****

**در اینکس قدیمی به عنوان مثال اگر عبارت فوتبال فرهاد را سرچ کنیم با توجه به اینکه idf کلمه فرهاد بیشتر است چون نادر تر است اسنادی بازگردانده می‌شوند که وزن فرهاد در آنها بیشتر است یا تعداد تکرار بیشتری از فرهاد دارند با مربوط تر به فرهاد اند**

**در ایندکس جدید ارزش کلمه فرهاد و مربی در کوئری یکسان است در نتیجه اسنادی که برمیگردند وزن این دو کلمه برابر حتی به خاطر بیشتر بودن تکرار کلمه مربی در اسناد معمولا اسناد با کلمه مربی بیشتر بازگردانده خواهند شد.**

**0.005855083465576172**

**Title: باشگاه استقلال: ابتلای مجیدی به کرونا صحت ندارد**

**Score: 0.17370253502851155**

**Link: https://www.farsnews.ir/news/14001027000699/باشگاه-استقلال-ابتلای-مجیدی-به-کرونا-صحت-ندارد**

**Content:**

**به گزارش خبرگزاری فارس، با اعلام باشگاه استقلال، تست کرونای فرهاد مجیدی منفی است و خبر ابتلای سرمربی استقلال به ویروس کرونا صحت ندارد. این در حالی است که سرمربی استقلال امروز در نشست خبری پیش از بازی با پیکان شرکت نکرده است. انتهای پیام/**

**در اینکدس اول با سرچ عبارت فرهاد فوتبال چون عملا idf کلمه فرهاد بیشتر است پس حضور کلمه فرهاد در سند مهمتر است و حضور دارد.**

**0.0045969486236572266**

**Title: استیلی سرپرست تیم ملی فوتبال شد**

**Score: 0.30669293102068246**

**Link: https://www.farsnews.ir/news/14001204000668/استیلی-سرپرست-تیم-ملی-فوتبال-شد**

**Content:**

**به گزارش خبرنگار ورزشی خبرگزاری فارس، حمید استیلی بازیکن اسبق تیم ملی فوتبال ایران به عنوان سرپرست تیم ملی فوتبال انتخاب شد. قرار است وی به زودی امور سرپرستی تیم ملی را بر عهده دارد. پیش از این مجتبی خورشیدی سرپرستی تیم ملی فوتبال را بر عهده داشت. انتهای پیام/**

**در اینکدس جدید چون تعداد سرچ کلمه فوتبال بیشتر بود و عملا وزن فرهاد به خاطر حذف idf کم شده بود عملا در سند برگردانده شده کلمه فرهاد وجود ندارد.**