

# گزارش فاز دوم پروژه بازیابی اطلاعات

نام و نام خواندگی: شیرین عبادی

شماره دانشجویی: ۹۷۳۴۰۱۶

۱.  
پرسमान ساده تک‌کلمه‌ای  
کوثری: «آمریکا»

مرتبط	جمله مربوطه	عنوان	لینک	شماره
تقریباً مرتبط است. طول آن نسبتاً کوتاه و کلمه آمریکا و تکرار آن در سند مشهود است.	ویزای یک نفر برای این سفر صادر نشود. تیم به آمریکا نمی‌رود.	ویزای 3 عضو کاروان گشتی... ی به آمریکا لغو می‌شود؟	<a href="#">لینک خبر</a>	۱
این خبر چنان مرتبط نیست چرا که علیرغم اینکه چندین بار کلمه آمریکا در آن تکرار شده است اما راجع به آن نیست.	ما قرار نیست مانند آمریکا باشیم.	باید با ارتقای قانون مال... به صنعت کشور بازگردانیم	<a href="#">لینک خبر</a>	۲
این خبر هم تقریباً مرتبط است چرا که تا حدودی راجع به تحریم‌های آمریکا است.	باید در مذاکرات پیش رو تضامینی را از آمریکا گرفت.	تضمین خواستن چین و روسیه... د رویکرد ایران صحیح است	<a href="#">لینک خبر</a>	۳
این خبر تقریباً مرتبط است.	علاوه بر اظهارات ضد آمریکایی دبیر، او گرین کارت آمریکا را دارد.	اظهارات ضدآمریکایی دبیر و احتمال عدم سفر به آمریکا	<a href="#">لینک خبر</a>	۴
این خبر تقریباً مرتبط است.	از عملیات مقتدرانه سپاه مقابل آمریکا تشکر کردند.	تقدیر مردم از مقابله سپاه با دزدی دریایی آمریکا در تجمع 13 آبان	<a href="#">لینک خبر</a>	۵

نتایج حاصل از champion list

این خبر تقریباً مرتبط است چرا که تا حدودی راجع به تحریم‌های آمریکا است.	باید در مذاکرات پیش رو تضامینی را از آمریکا گرفت.	تضمین خواستن چین و روسیه... د رویکرد ایران صحیح است	<a href="#">لینک خبر</a>	۱
کلمات آمریکا در این سند با فرکانس بالایی مشاهده می‌شود و مرتبط است اما بیشتر راجع به ارتباط ایران	ایزدی معتقد است آمریکایی‌ها برای زیستن نیاز به بحران دارند	انقلاب، ثبات، امنیت و اقتدار نظامی دشمنان دیگر جرات نمی‌کنند ایران را تهدید نظامی کنند	<a href="#">لینک خبر</a>	۲

				و امریکا است و بین اسنادی که داریم تقریباً مرتبط است.
۳	<a href="#">لینک خبر</a>	آمریکا دولت‌های متحد خود را در زمان اضطراب تنها می‌گذارد.	و عده‌های آمریکا برای حمایت از آنها دروغ است.	این سند نیز با تکرار بالایی کلمه آمریکا را شامل می‌شود و مرتبط است.
۴	<a href="#">لینک خبر</a>	درخواست کشتی آمریکا برای تجدیدنظر/مبارزه دوستانه به نفع هر ۲ کشور است	فدراسیون کشتی آمریکا از لغو این دیدار دوستانه خبر داد	مرتبط است.
۵	<a href="#">لینک خبر</a>	فقط عوامل آمریکا در داخل کشور از مقابله ایران با دزدان دریایی ناراحتند	این افراد عوامل آمریکا در داخل کشور هستند	در این سند کلمه آمریکا با فرکانس بالایی آمده‌اند و به‌طور کلی نیز مرتبط به آمریکا است. طول این سند نسبتاً بلند است.

**پرسیمان چند کلمه‌ای متداول**  
**کوثری: «ورزش بانوان ایران»**

۱	<a href="#">لینک خبر</a>	پاسخ مثبت آجورلو به وزیر ورزش برای تشکیل تیم بانوان استقلال	وزیر ورزش و جوانان کشورمان در مراسم جشن و اگذاری سهام سرخابی‌ها، از انتصاب مسئول فوتبال بانوان باشگاه استقلال خبر داد.	سند طول کوتاهی دارد و بعلاوه کلمه بانوان و ورزش در آن نسبت به طولش تکرار زیادی دارند. در کل مرتبط است چرا که مربوط به شاخه‌ای از ورزش زنان است
۲	<a href="#">لینک خبر</a>	سجادی: تحول بزرگ در عرصه ورزش بانوان در پیش است	در دولت سیزدهم اداره ورزش بانوان را به خودشان می‌سپاریم	کاملاً مرتبط است. در این سند بارها کلمه ورزش بانوان تکرار شده است.
۳	<a href="#">لینک خبر</a>	حضور مسئولین ورزش و نمای... ن تیم ملی فوتبال بانوان	برای اولین در تاریخ فوتبال بانوان ایران اتفاق افتاده است، شهاب‌الدین عزیزی خادم، رییس فدراسیون میزبان مسئولین ورزش کشور و نمایندگان	مرتبط است. در این سند کلمه بانوان تکرار شده است و کلمه ورزش نیز در آن مشهود است. به‌طور کلی مانند سند یک زیرشاخه‌ای از ورزش

			مجلس شورای اسلامی در تمرین تیم ملی فوتبال بانوان خواهد بود.	بانوان محسوب می‌شود و مرتبط است.
۴	<a href="#">لینک خبر</a>	"کاظمی‌پور: فدراسیون ها نتیجه توجه به ورزش بانوان را خواهند دید"	موضوعات مهم مرتبط با ورزش بانوان با هم افزایی و هماهنگی، نتایج بهتری در راستای توسعه و تعالی ورزش بانوان حاصل شود.	مرتبط است. این سند نسبتاً طولانی با تکرار بالایی شامل ورزش بانوان است.
۵	<a href="#">لینک خبر</a>	"۲ رشته تویی بانوان به لیست اعزامی به بازی‌های آسیایی اضافه شدند"	در این بین حمایت و پشتیبانی از ورزش بانوان نیز در دستور کار است بطوریکه اولویت حضور فرهنگی بانوان در این رویدادها هم مورد توجه و تاکید قرار گرفته است.	مرتبط است.

نتایج حاصل از champion list

۱	<a href="#">لینک خبر</a>	سجادی: تحول بزرگ در عرصه ورزش بانوان در پیش است	در دولت سیزدهم اداره ورزش بانوان را به خودشان می‌سپاریم.	کاملاً مرتبط است.
۲	<a href="#">لینک خبر</a>	"کاظمی‌پور: فدراسیون ها نتیجه توجه به ورزش بانوان را خواهند دید"	موضوعات مهم مرتبط با ورزش بانوان با هم افزایی و هماهنگی، نتایج بهتری در راستای توسعه و تعالی ورزش بانوان حاصل شود.	کاملاً مرتبط است.
۳	<a href="#">لینک خبر</a>	"۲ رشته تویی بانوان به لیست اعزامی به بازی‌های آسیایی اضافه شدند"	در این بین حمایت و پشتیبانی از ورزش بانوان نیز در دستور کار است بطوریکه اولویت حضور فرهنگی بانوان در این	مرتبط است.

			رویدادها هم مورد توجه و تاکید قرار گرفته است.	
۴	<a href="#">لینک خبر</a>	پاسخ مثبت آجورلو به وزیر ورزش برای تشکیل تیم بانوان استقلال	وزیر ورزش و جوانان کشورمان در مراسم جشن واگذاری سهام سرخابی‌ها، از انتصاب مسئول فوتبال بانوان باشگاه استقلال خبر داد.	مرتبط است.
۵	<a href="#">لینک خبر</a>	"واکنش نماینده مجلس به بحث استفاده از جادوگر در فدراسیون فوتبال"	از آنجایی هم که سلامت یک جامعه در ید ورزش و نشاطی است که از ورزش نشأت می‌گیرد، نباید نقش بانوان را در آن نادیده گرفت.	مرتبط است

#### نتایج حاصل از فاز یک

۱	<a href="#">لینک خبر</a>	پاسخ مثبت آجورلو به وزیر ورزش برای تشکیل تیم بانوان استقلال	وزیر ورزش و جوانان کشورمان در مراسم جشن واگذاری سهام سرخابی‌ها، از انتصاب مسئول فوتبال بانوان باشگاه استقلال خبر داد.	کاملاً مرتبط است.
۲	<a href="#">لینک خبر</a>	بار دیگر فدراسیون فوتبال علیه مجلس/نامه نگاری غیرواقعی به فیفا برای تصویب اساسنامه	-	این سند به بحث ورزش مرتبط است اما نه ورزش زنان.
۳	<a href="#">لینک خبر</a>	"در نسبت انقلاب اسلامی و ورزش بانوان/ زنان ایرانی چگونه کانون توجهات جهانیان شدند؟"	پرچمداری زنان در چند دوره از المپیک نشان از توجه ایران اسلامی به ورزش بانوان داشت.	مرتبط است.
۴	<a href="#">لینک خبر</a>	تجلیل از دختر تاریخ ساز تنیس با پاداش هزار دلاری	مراسم تجلیل و بدرقه مشکات الزهرا صفی بانوی تاریخ ساز ورزش ایران با حضور سید رضا	مرتبط است. در کل می‌توان آن را زیرشاخه‌ای از ورزش بانوان دانست.

			صالحی امیری رئیس کمیته ملی المپیک، مریم کاظم پور معاون توسعه ورزش بانوان وزارت ورزش و جوانان ....	
۵	<a href="#">لینک خبر</a>	"۲ رشته توپی بانوان به لیست اعزامی به بازی‌های آسیایی اضافه شدند"	در این بین حمایت و پشتیبانی از ورزش بانوان نیز در دستور کار است بطوریکه اولویت حضور فرهنگی بانوان در این رویدادها هم مورد توجه و تاکید قرار گرفته است.	مرتبط است.

همان‌طور که از نتایج مشخص است در روش tf-idf سعی می‌کنیم اسنادی که شامل هر دو کلمه هستند و با سنجش نسبت تکرارشان به نسب اندازه کل سند بازگردانیم. در روش ایندکس معمولی فاز یک بیشتر برایمان مهم است که تمامی کلمات در سند موجود باشند و به تکرار آن‌ها یا طول سند توجهی نمی‌کنیم. بنابراین ممکن است سندی که تنها یکبار ورزش بانوان در آن آمده اما ربطی به این موضوع ندارد را بازگردانیم (مانند نتیجه شماره ۲). به‌طور کلی نتایج هر دو روش در این کوئری تقریباً مشابهند و مرتبط هم هستند.

#### پرسمان تک‌کلمه‌ای پیچیده

کوئری: صهیونیست

۱	<a href="#">لینک خبر</a>	"محکومیت حضور کاروان ورزشی رژیم صهیونیستی در امارات+عکس"	به همین خاطر این سنادیکا به حضور ورزشکاران <b>صهیونیست</b> واکنش نشان داده است.	مرتبط است
۲	<a href="#">لینک خبر</a>	"اتفاق عجیب در ادامه عادی سازی روابط رژیم اشغالگر با جهان اسلام/ مهدوی کیا کنار مربی صهیونیست؟"	حاضر نشدند در کنار یک <b>صهیونیست</b> بازی کنند.	مرتبط است.
۳	<a href="#">لینک خبر</a>	"اقدام حمایتی عمان از فلسطین/ سرود و پرچم رژیم اشغالگر بایکوت شد+عکس"	غال زوکرمان ورزشکار بانوی <b>صهیونیست</b> در این مسابقات جهانی مدال طلا گرفت.	مرتبط است.
۴	<a href="#">لینک خبر</a>	"فیفا بگوید چرا کشورهای متجاوز به یمن را محروم	چرا رژیم <b>صهیونیستی</b> محروم نمی‌شود.	این سند بطور مستقیم به صهیونیست مرتبط نیست اما کلمه

صهیونیست در آن بکار رفته است و اشاره‌ای به آن شده است. تقریباً مرتبط.		نمی‌کند؟"		
	۵			

نتایج ۵ و ۴ مشابه هستند. علیرغم اینکه دو سند متفاوت با شماره سند (۹۸۲ و ۷۳۱۷) هستند اما هر دو یک سند یکسانند.

نتایج حاصل از champion list

مرتبط است	به همین خاطر این سنادیکا به حضور ورزشکاران <b>صهیونیست</b> واکنش نشان داده است.	"محکومیت حضور کاروان ورزشی رژیم صهیونیستی در امارات+عکس"	<a href="#">لینک خبر</a>	۱
مرتبط است.	حاضر نشدند در کنار یک <b>صهیونیست</b> بازی کنند.	"اتفاق عجیب در ادامه عادی سازی روابط رژیم اشغالگر با جهان اسلام/ مهدوی کیا کنار مربی صهیونیست؟"	<a href="#">لینک خبر</a>	۲
مرتبط است.	غال زوکرمان ورزشکار بانوی <b>صهیونیست</b> در این مسابقات جهانی مدال طلا گرفت.	"اقدام حمایتی عمان از فلسطین/ سرود و پرچم رژیم اشغالگر بایکوت شد+عکس"	<a href="#">لینک خبر</a>	۳
این سند بطور مستقیم به صهیونیست مرتبط نیست اما کلمه صهیونیست در آن بکار رفته است و اشاره‌ای به آن شده است. تقریباً مرتبط.	چرا رژیم <b>صهیونیستی</b> محروم نمی‌شود.	"فیفا بگوید چرا کشورهای متجاوز به یمن را محروم نمی‌کند؟"	<a href="#">لینک خبر</a>	۴

پرسمان چندکلمه‌ای پیچیده

کوثری: طرح منع تردد»

مرتبط است	طرح منع تردد شبانه غیر از جریمه کردن مردم و افزایش هزینه‌های زندگی‌شان، چه تأثیری در کاهش کرونا داشته است؟	<a href="#">لینک خبر</a>	۱
مرتبط است	منع تردد شبانه درحالی هنوز در کشور در حال اجرا است که اعمال این محدودیت از سه منظر قابل بررسی است.	<a href="#">لینک خبر</a>	۲
این خبر به کوثری چندان مرتبط نیست. علیرغم آنکه کلمه منع تردد یکبار در آن دیده می‌شود اما سند مربوط به موضوع دیگری است.	از رئیس جمهور برای لغو منع تردهای شبانه تشکر کرد.	<a href="#">لینک خبر</a>	۳
این سند کلمه منع را ندارد اما ممنوعیت در آن مشهود است. بعلاوه طول آن نسبت به سند بالا بیشتر است. سند نیز مرتبط است.	در توضیح آخرین وضعیت طرح ممنوعیت تردد شبانه، گفت: ....	<a href="#">لینک خبر</a>	۴
مرتبط است.	درباره مصوبه منع تردد شبانه نیز در ستاد ملی کرونا بحث شد ولی لغو آن تصویب نشد.	<a href="#">لینک خبر</a>	۵

نتایج حاصل از champion list

مرتبط است	طرح منع تردد شبانه	"منع تردد شبانه غیر	<a href="#">لینک خبر</a>	۱
-----------	--------------------	---------------------	--------------------------	---



		غیر از جریمه کردن مردم و افزایش هزینه‌های زندگی‌شان، چه تأثیری در کاهش کرونا داشته است؟	از جریمه کردن مردم، چه تأثیری در کرونا داشته است؟	
مرتبط است	منع تردد شبانه درحالی هنوز در کشور در حال اجرا است که اعمال این محدودیت از سه منظر قابل بررسی است.	"خضریان: با اعمال منع تردد شبانه، شهرهای بزرگ برای گشت و گذار ثروتمندان قُرق شده است"	<a href="#">لینک خبر</a>	۲
این خبر به کوثری چندان مرتبط نیست. علیرغم آنکه کلمه منع تردد یکبار در آن دیده می‌شود اما سند مربوط به موضوع دیگری است.	از رئیس جمهور برای لغو منع تردهای شبانه تشکر کرد.	"نماینده اهواز: میزان سرقت در استان خوزستان افزایش یافته است"	<a href="#">لینک خبر</a>	۳
این سند کلمه منع را ندارد اما ممنوعیت در آن مشهود است. بعلاوه طول آن نسبت به سند بالا بیشتر است. سند نیز مرتبط است.	در توضیح آخرین وضعیت طرح ممنوعیت تردد شبانه، گفت: ....	وحیدی: ممنوعیت تردد شبانه کماکان پابرجاست/ لایحه جامع انتخابات را به مجلس ارسال می‌کنیم	<a href="#">لینک خبر</a>	۴
مرتبط است.	درباره مصوبه منع تردد شبانه نیز در ستاد ملی کرونا بحث شد ولی لغو آن تصویب نشد.	وزیر کشور: با افزایش تخریق 2 دز واکسن محدودیت‌های تردد کاهش می‌یابد	<a href="#">لینک خبر</a>	۵

در کل می‌توان گفت champion list در مجموع و در یکی از سندها با نتایج معمولی اختلاف دارد و نتیجه champion list مرتبط‌تر است.

نتیجه حاصل از فاز یک

تقریباً مرتبط. یک سند طولانی که در بخشی از آن به منع تردد اشاره می‌شود.	منع تردد شبانه و درآمدزایی آن برای دولت گفت..	اعمال محدودیت شبانه برای دولت منبع درآمد نبوده/ تصویب دستورالعمل قرنطینه هوشمند در ستاد ملی کرونا	<a href="#">لینک خبر</a>	۱
---	---	---	--------------------------	---

مرتبط است	منع تردد شبانه در حالی هنوز در کشور در حال اجرا است که اعمال این محدودیت از سه منظر قابل بررسی است.	"خضریان: با اعمال منع تردد شبانه، شهرهای بزرگ برای گشت و گذار ثروتمندان فُرَق شده است"	<a href="#">لینک خبر</a>	۲
مرتبط است	طرح منع تردد شبانه غیر از جریمه کردن مردم و افزایش هزینه‌های زندگی‌شان، چه تأثیری در کاهش کرونا داشته است؟	"منع تردد شبانه غیر از جریمه کردن مردم، چه تأثیری در کرونا داشته است؟"	<a href="#">لینک خبر</a>	۳
غیر مرتبط. تنها کلمه تردد و منع را در خود دارد و به‌طور کل غیر مرتبط است.	-	فنون‌زاده: ذوالفقار نسب به جای تلویزیون باید در VIP ورزشگاه بنشیند/ درویش هوادار پرسپولیس بود	<a href="#">لینک خبر</a>	۴
یک سند بسیار طولانی که هیچ ارتباطی به کوئری ندارد و تنها کلمه منع و تردد در آن چند بار تکرار شده‌اند.	-	قرائت گزارش بررسی عملکرد حسن روحانی در ستاد کرونا/ مردم از عملکرد دولت سابق علیه کرونا ناراضی هستند	<a href="#">لینک خبر</a>	۵

در مورد این کوئری مشاهده می‌شود که روش **tf-idf** بسیار بهتر از روش ایندکس معمولی فاز یک عمل می‌کند. با در نظر گرفتن طول اسناد و وزن کلمه در سند می‌توان به نتایج بهتری رسید و اسکورینگ بهتری داشت نسبت به حالتی که تنها تعداد تکرار کلمات در سند را در نظر بگیریم. در حالت ایندکس معمولی اسنادی بازگردانده شدند که در یک نوشته طولانی کلمات کوئری به‌صورت پراکنده حضور داشتند اما به هیچ وجه مرتبط به کوئری نبودند. یکی از علت‌های آن صرفاً در نظر گرفتن این نکته است که اسکورینگ را بر مبنای حضور بیشترین کلمات کوئری در سند انجام دهیم، و یا تکرار آن در سند را مد نظر قرار دهیم. در مورد **champion list** و حالت شاخص معمولی نیز می‌توان گفت که نتایج حاصل از **champion list** و حاصل شاخص معمولی چندان تفاوتی با هم ندارند اما به‌طور کلی **champion list** هم سریعتر است و هم در مواردی بهتر عمل می‌کند. عملکرد ایندکسینگ فاز یک چندان مطلوب نیست.