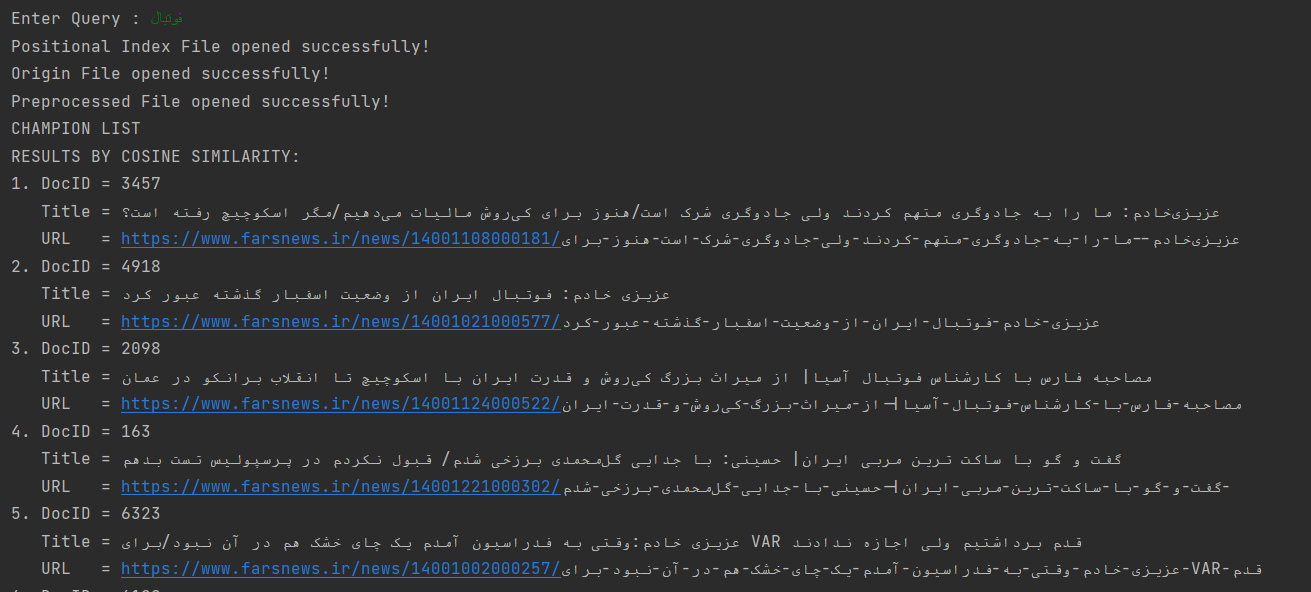
گزارش پروژه بازیابی اطلاعات

فاز 2- بخش اول: مدل سازي اسناد در فضاي برداری



کوئری "فوتبال" را وارد میکنیم، ابتدا پاسخ ها با شباهت کسینوسی نمایش داده میشود:



اولین پاسخ را بررسی میکنیم:

همانطور که مشاهده میشود 76 بار در این مطلب کلمه فوتبال امده و موضوع مطلب نیز با کوئری وارد شده مرتبط است.

سپس پاسخ های شناسایی شده با ضریب ژاکارد را بررسی میکنیم:



اولین پاسخ را تحلیل میکنیم:

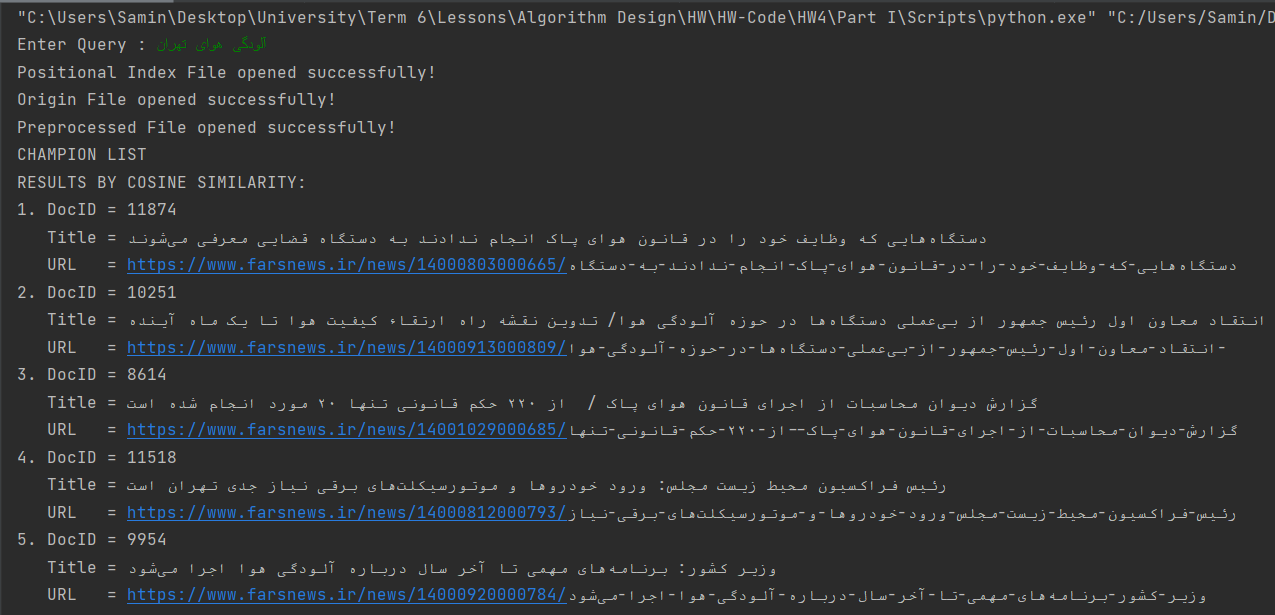


همانطور که دیده میشود 31 بار کوئری در این مطلب اورده شده و محتوای مرتبطی نیز دارد.



عبارت "آلودگی هوای تهران" را وارد میکنیم:

شباهت کسینوسی



اولین پاسخ را بررسی میکنیم:

همانطور که مشاهده میشود عبارت خواسته شده در این مطلب موجود بود و 3 بار تکرار شده بود و محتوای مطلب نیز مرتبط است.

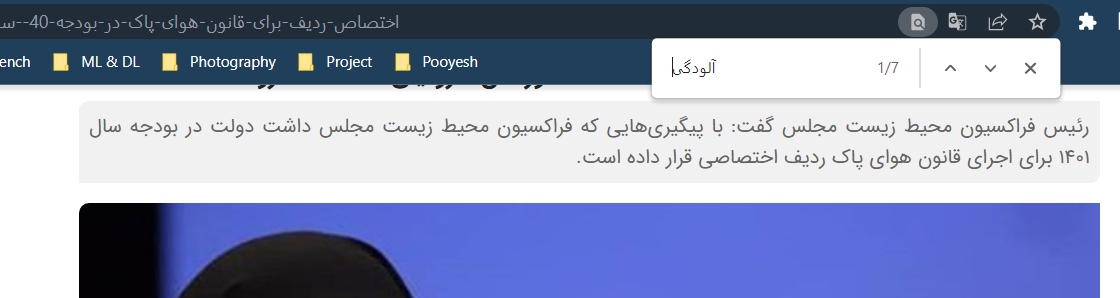
ژاکارد:



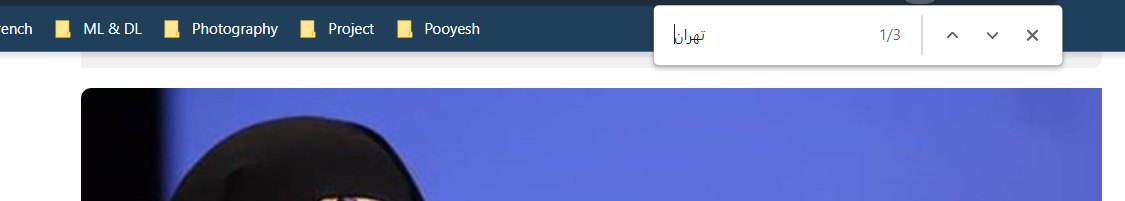
اولین پاسخ را بررسی میکنیم:



در این مطلب عبارت مستقلا یافت نشد که قابل قبول است زیرا ما در این بخش از پروژه امکان بررسی عبارت را پیاده سازی نکرده بودیم. اما محتوای مطلب مرتبط است و از طرفی کلمات را بصورت مجزا نیز دارد.



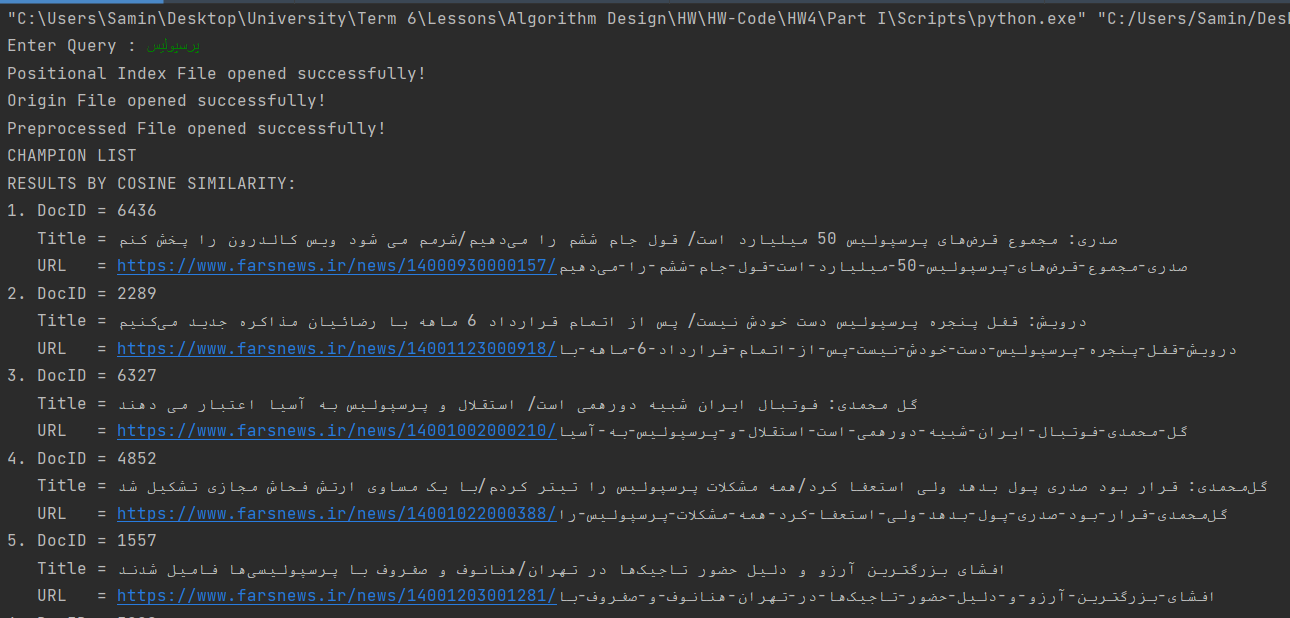






کلمه پرسپولیس را جستجو میکنیم:

بر اساس شباهت کسینوسی:



اولین پاسخ را بررسی میکنیم:



همانطور که مشاهده میشود 102 مورد پیدا شده و مطلب نیز مرتبط است.

ضریب شباهت ژاکارد:



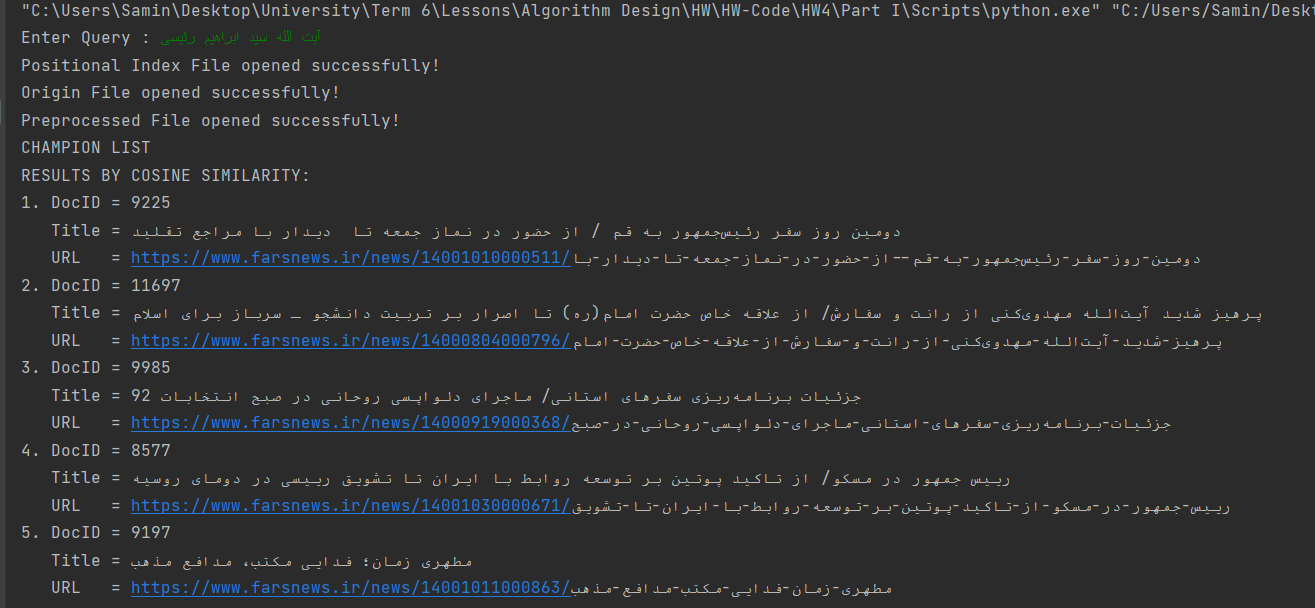
اولین پاسخ را بررسی میکنیم:

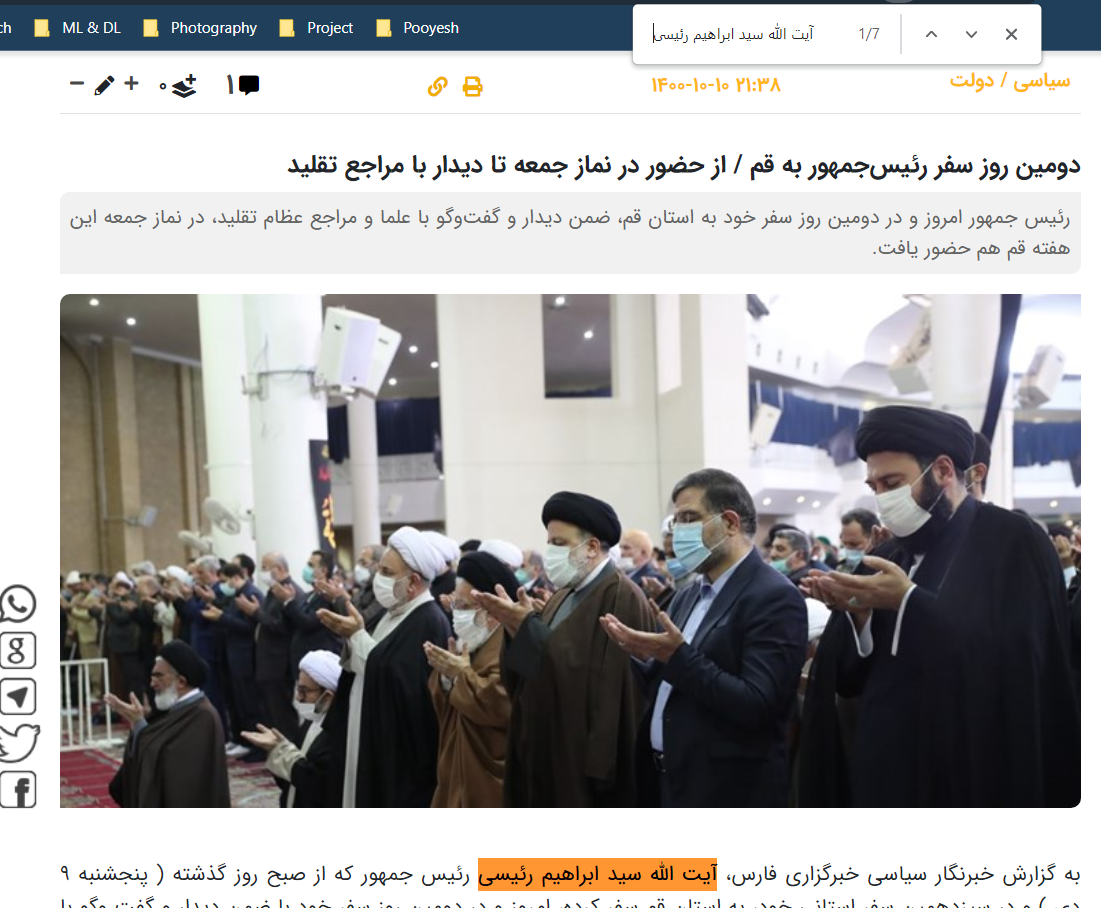
همانطور که مشاهده میشود 44 بار تکرار شده و مطلب نیز مرتبط است.



عبارت "آیت الله سید ابراهیم رییسی" وارد میکنیم:

بر اساس شباهت کسینوسی:

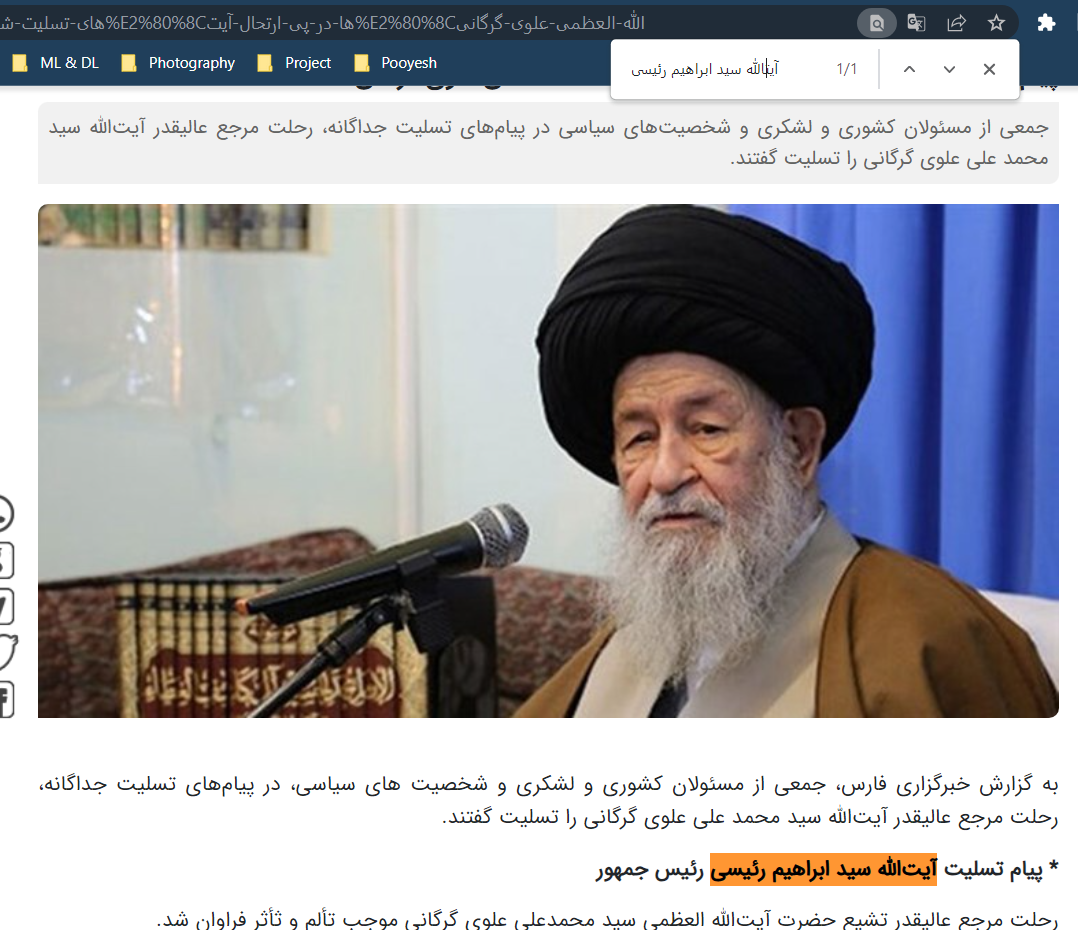


اولین پاسخ را بررسی میکنیم:

همانطور که مشاهده میشود عبارت 7 بار در مطلب امده و مرتبط نیز هست.

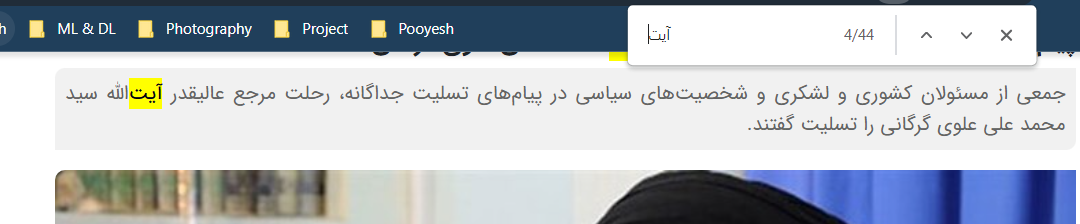
ضریب شباهت ژاکارد:



اولین پاسخ را بررسی میکنیم:

همانطور که مشاهده میشود عبارت در مطلب آورده شده است( علت تفاوت عبارت جستجو شده استفاده از نیم فاصله به جای فصله در مطلب اصلی بود که در موتور جستجوی ما تفاوتی ایجاد نمیکند ولی برای جستجو در صفحه وب مشکل زاست)

از طرفی عبارت تکرار زیادی ندارد که باز هم به مشکل عدم پیاده سازی عبارات برمیگردیم ولی کلمات همگی به مراتب تکرار شده اند.

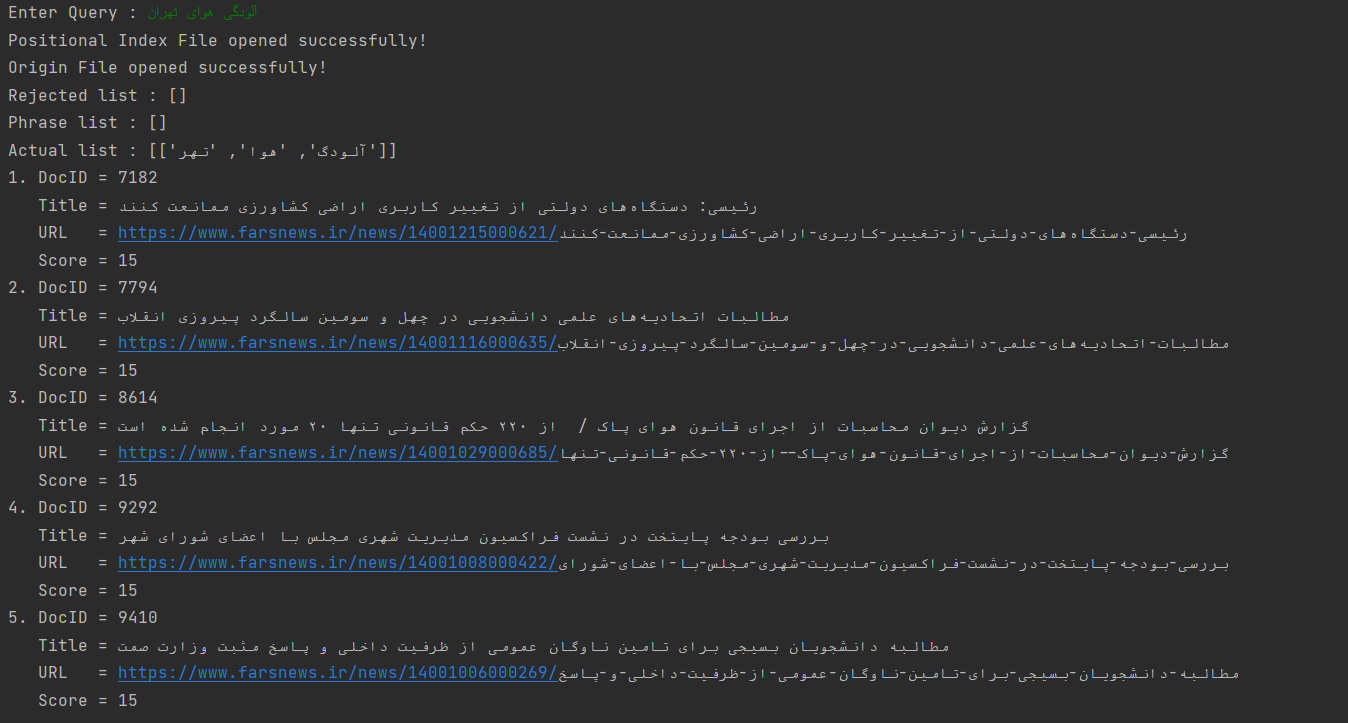




و به همین ترتیب.



مورد ب) آلودگی هوای تهران



اولین پاسخ را بررسی میکنیم:

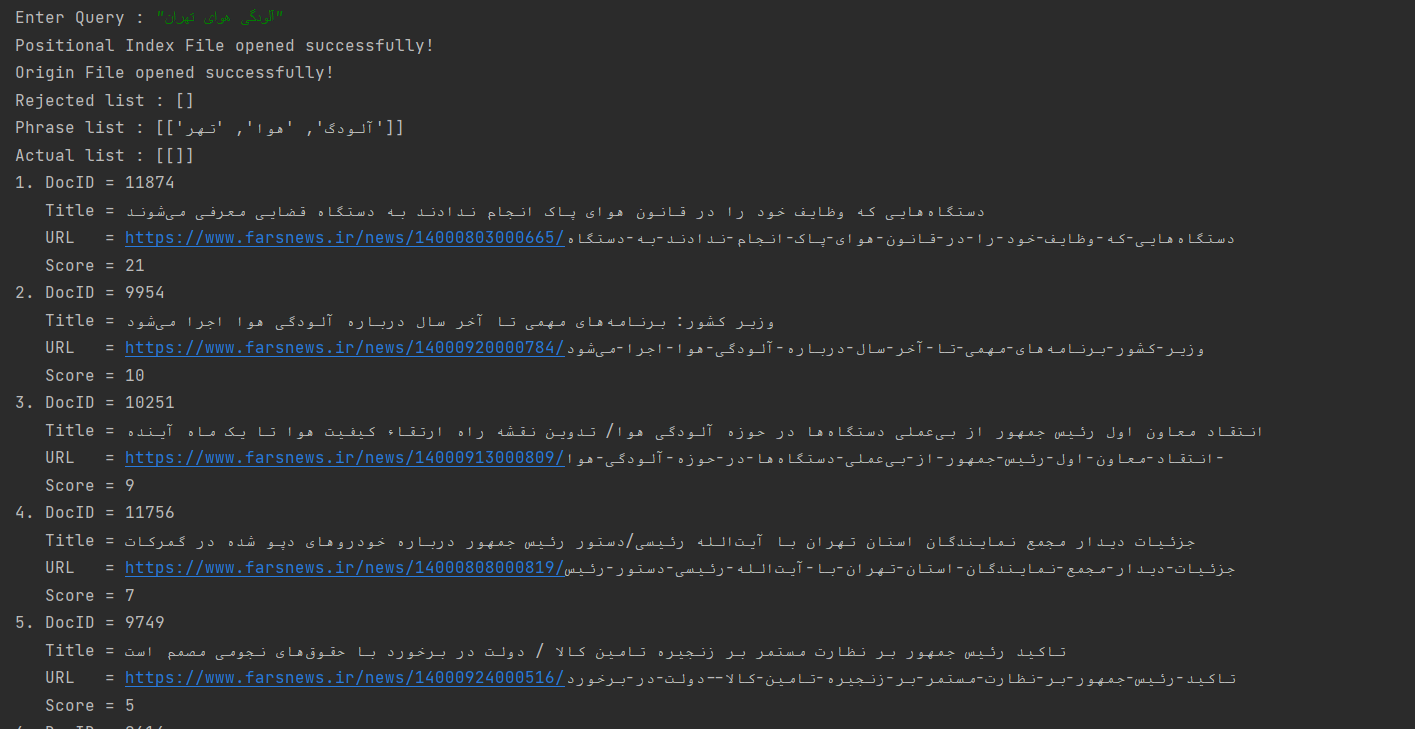


عبارت مستقلا یافت نشده که ما هم آن را بصورت عبارت جستجو نکرده بودیم ولی کلمات بصورت مجزا وجود دارند.



و مشاهده میشود مطلب خیلی مرتبط نیست.

حالا بصورت عبارت هم جستجو میکنیم:



مشاهده میکنیم که جواب های بهتری گرفتیم و از طرفی شباهت بین پاسخ های این بخش مشابه فاز 2 هم هست.

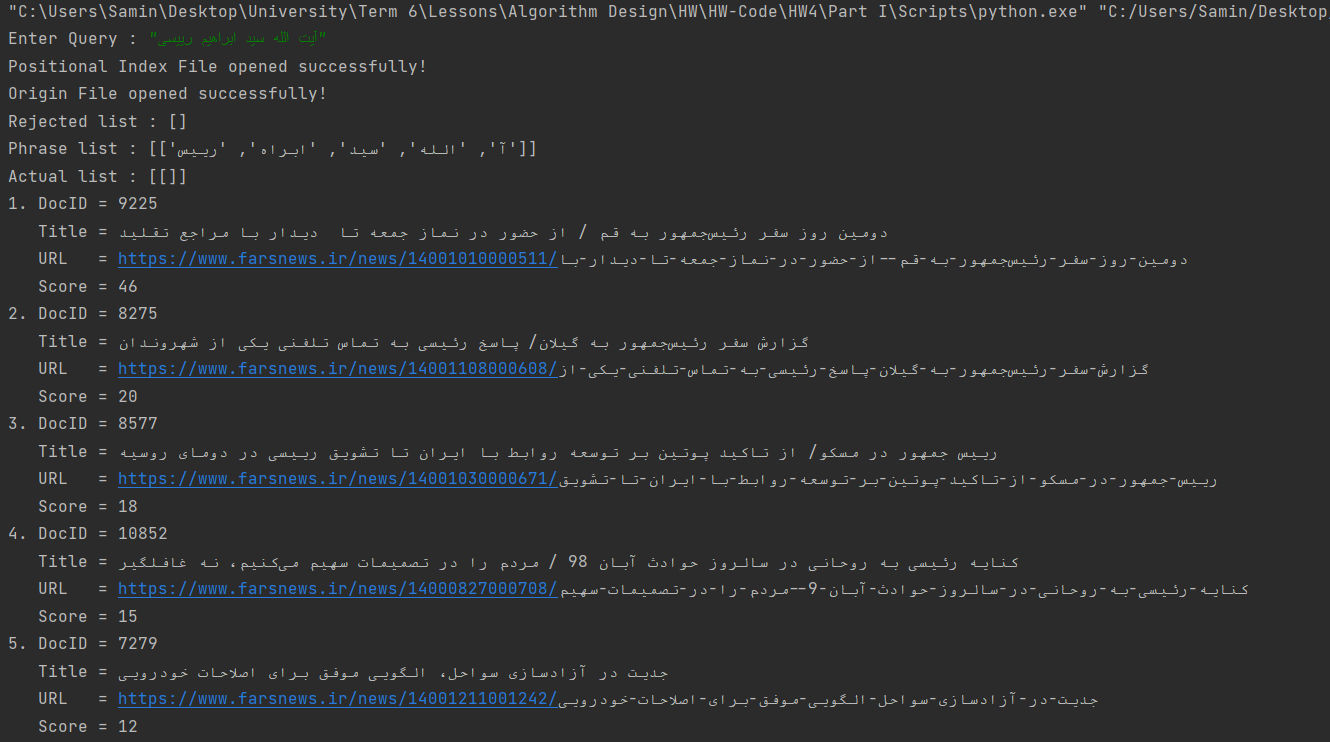
اولین پاسخ را بررسی میکنیم:



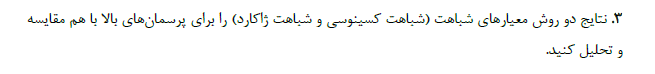
همانطور که مشاهده میشود پاسخ این بخش و بخش قبل در شباهت کسینوسی مشابه بوده، درصورتی که ما در شباهت کسینوسی جستجوی عبارات را پیاده سازی نکرده بودیم ولی پاسخی همانند آن گرفتیم درصورتی که پاسخ های ضریب شباهت ژاکارد چنین شرایطی نداشتند.

ت) آیت الله سید ابراهیم رییسی

عبارت را در فاز 1 جستجو میکنیم:



مشاهده میشود که مجددا پاسخی مشابه بخش شباهت کسینوسی گرفتیم.



در هر بخش این مطلب توضیح داده شده ولی بصورت کلی میتوان گفت که شباهت کسینوسی توانایی بالاتری برای پاسخ گویی به عبارات دارد و از طرفی مطالب مرتبط تر با تعداد تکرار کلمه مستقل بیشتری را نتیجه داده است.

عملکرد ضریب ژاکارد در کلمات مستقل خوب بوده ولی باز هم در مقایسه با شباهت کسینوسی عملکرد خوبی نداشته است.