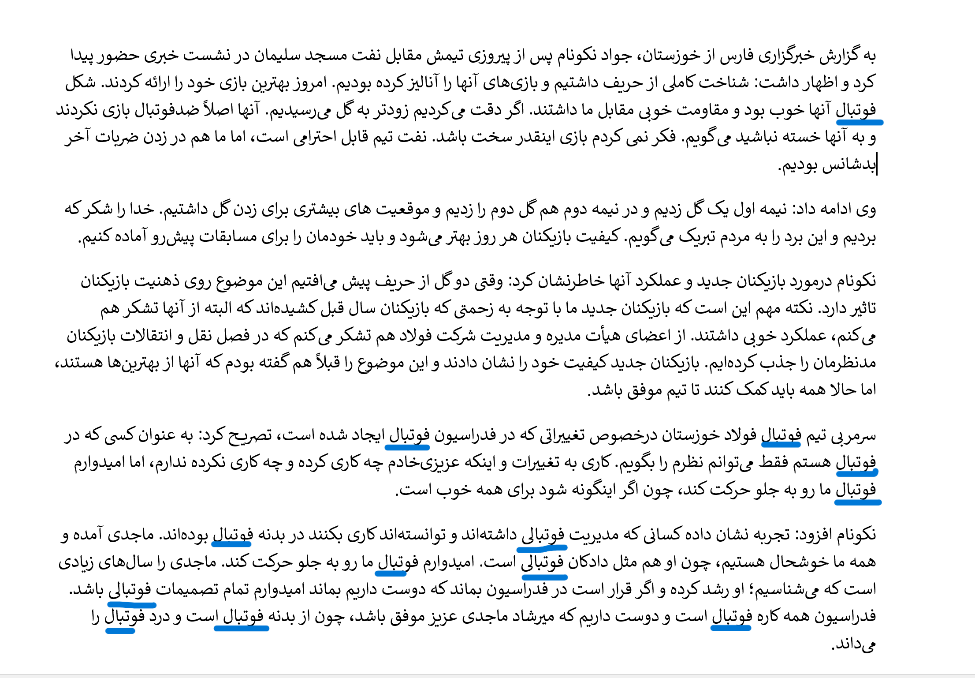
به نام خدا

**گزارش فاز دوم پروژه**

۱- الف) یک پرسمان از کلمات ساده و متداول تک کلمه ای:



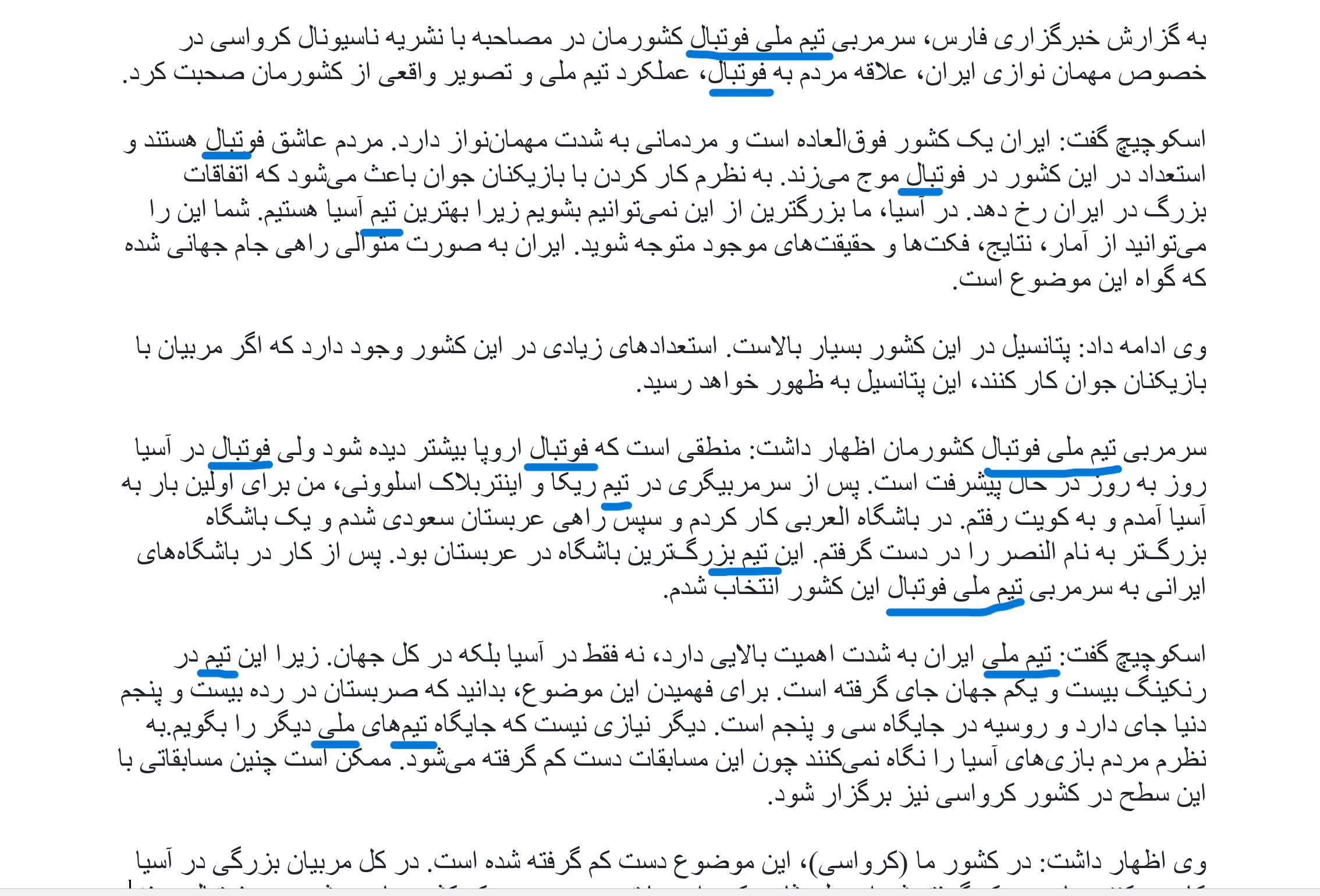
همانطور که از تیتر اخبار مشخص هست همگی مربوط به فوتبال می باشند. به عنوان مثال در تیتر خبر اول کلمه فوتبال مشاهده نمی شود اما اگر خود خبر را بخوانیم بارها کلمه فوتبال را میبینیم:



ب) یک پرسمان از عبارات ساده و متداول چندکلمه ای:



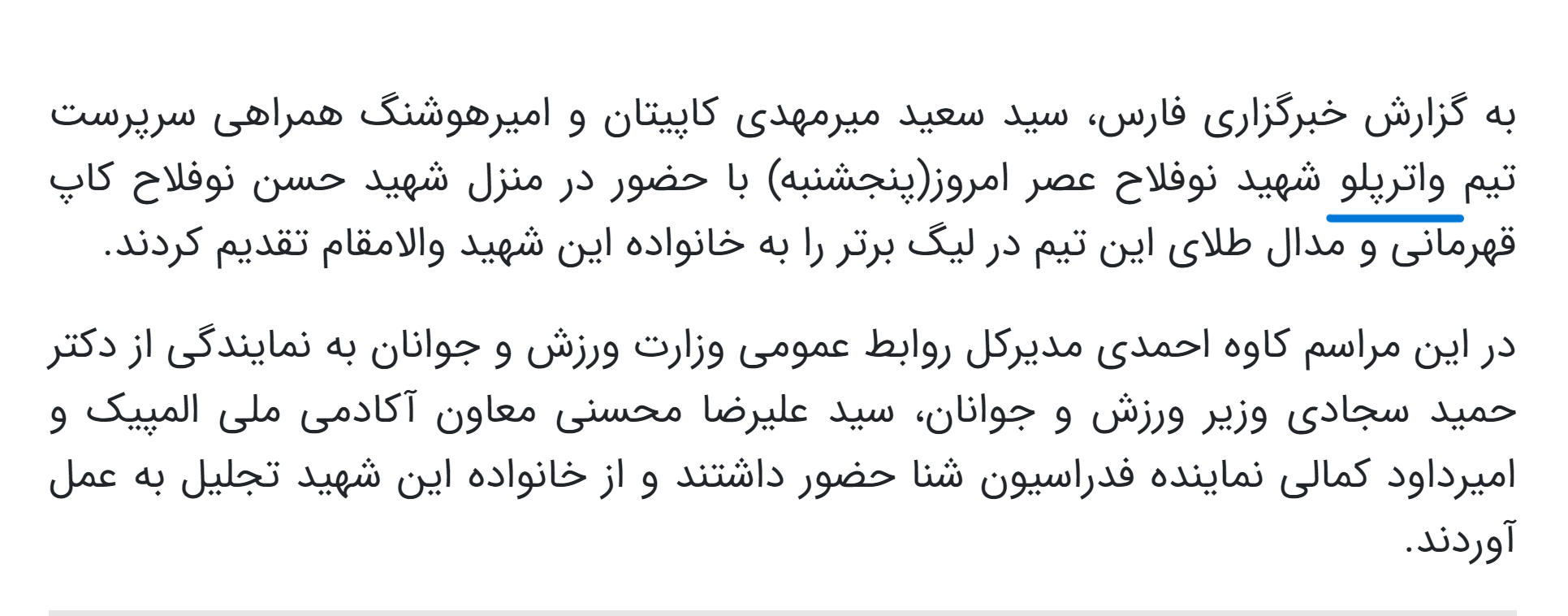
همانطور که از تیتر اخبار مشخص است همگی درباره ی تیم ملی فوتبال است. بخشی از متن خبر اول را مشاهده می کنیم:



پ) یک پرسمان دشوار و کم تکرار تک کلمه ای:



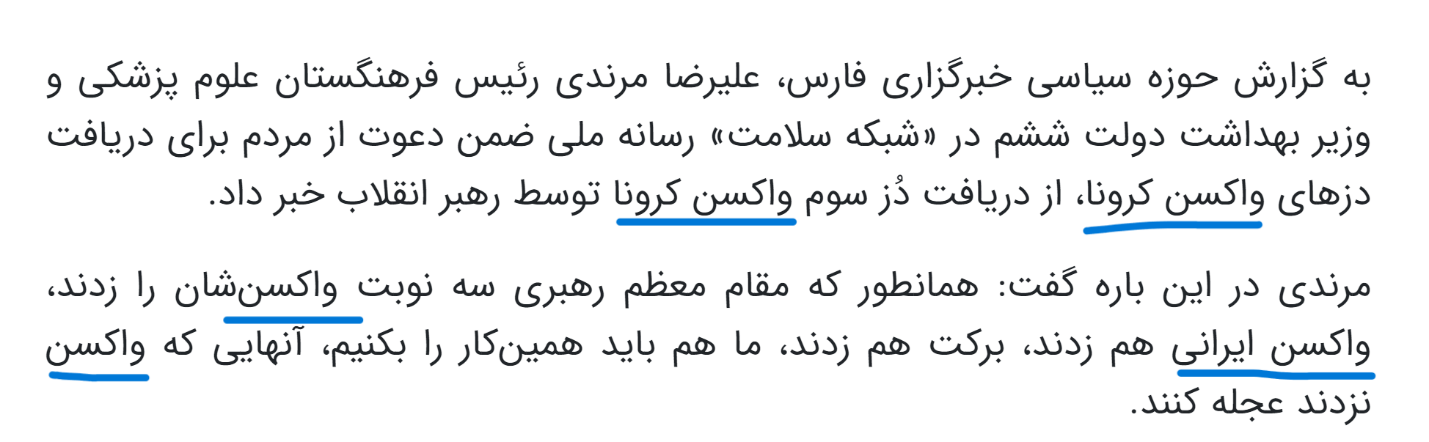
این کلمه کم تکرار است. همان طور که مشاهده می شود از تیتر اخبار مشخص است که مربوط به واترپلو می باشند. بخشی از متن خبر اول را مشاهده می کنیم:



ت) یک پرسمان دشوار و کم تکرار چندکلمه ای:



این عبارت کم تکرار است. همانطور که میبینیم نتایج به دست آمده تا حد زیادی مرتبط است. مثلا خبر اول که بالاترین امتیاز را دارد درباره این است که رهبر واکسن ایرانی زده است و خبر دوم درباره ی واکسن فخرا است که یک واکسن ایرانی است. بخشی از متن خبر اول را می بینیم:



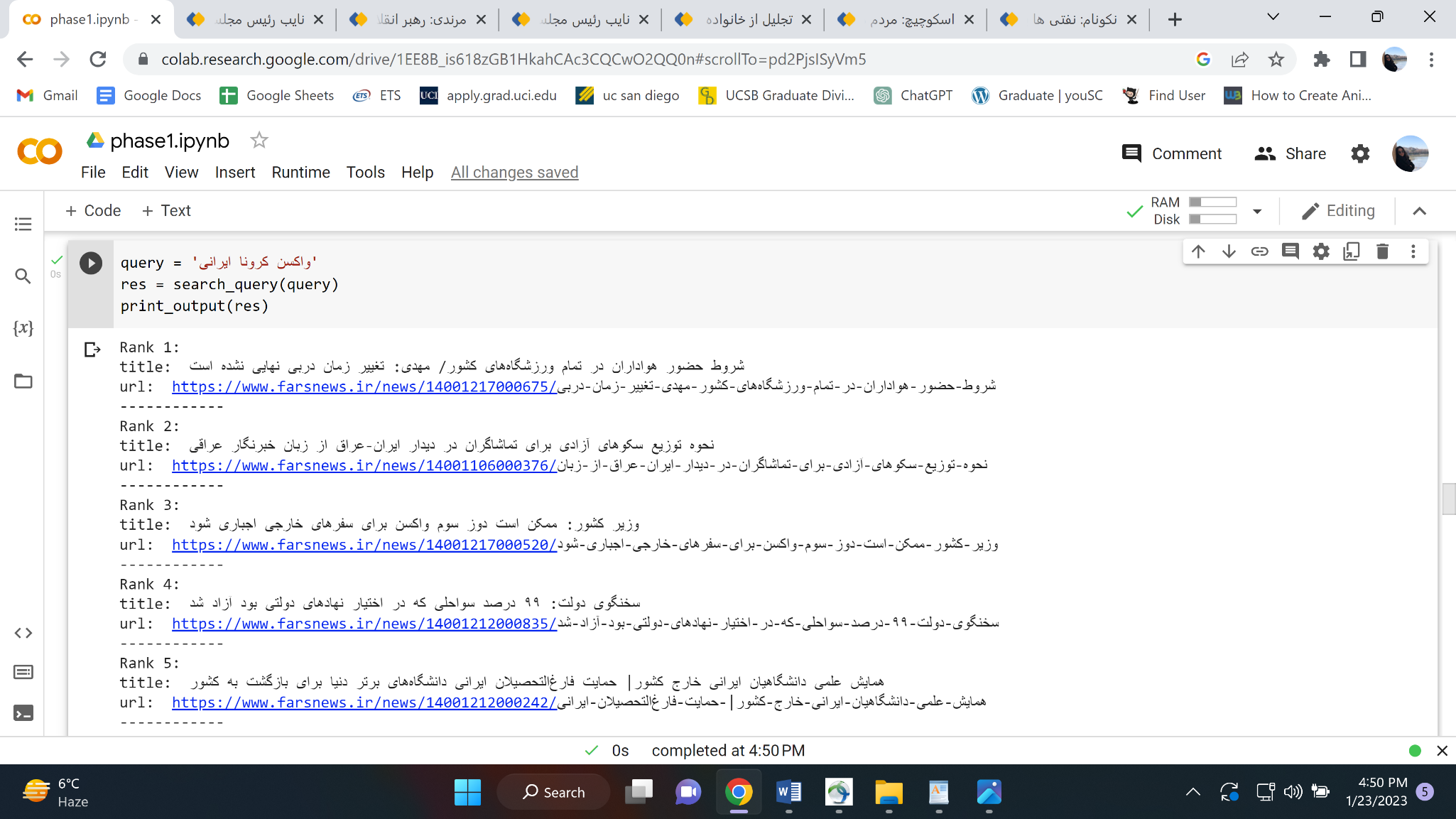
۲- نتایج قسمت ب و ت در فاز یک

ب)



همان طور که مشاهده می کنیم در برخی موارد نتایج مرتبط گرفتیم ولی اخبار اول و دوم صرفا به فوتبال و تیم مربوط بوده و درباره ی تیم ملی فوتبال نیست.

ت)



همان طور که مشاهده میکنیم نتایج تقریبا نامربوط اند.

نتایج بدست آمده برای پرسمان دشوار و کم تکرار، تقریبا نامرتبط است. در فاز یک ما اسنادی را که حاوی کلمات پرسمان هستند را بازیابی میکنیم و اسناد را بر اساس اینکه در کدام سند تعداد بیشتری از کلمات پرسمان آمده است مرتب میکنیم. برای مثال در پرسمان «واکسن کرونا ایرانی» اسنادی را که کلمات واکسن»، «کرونا» و «ایرانی را دارند در رتبه های بالاتر قرار میگیرند و اسنادی که بعضی از این کلمات را ندارند در رتبه های بندی قرار میگیرند.

در فاز دوم ما تعداد تکرار یک کلمه در یک سند و اهمیت آن کلمه در کل اسناد را با استفاده از tf-idf پیدا و اسناد می کنیم و با استفاده از شباهت کسینوسی اسناد را از نظر شباهت مقایسه می کنیم. پس هر کلمه یک وزن دارد و اسناد بر اساس وزن و اهمیت کلمات موجود در پرسمان بازیابی و رتبه بندی میشوند.

برای همین است که ما در فاز دوم نتایج دقیقتری گرفتیم و توانستیم به خوبی نتایج را رتبه دهی کنیم.