هزار خبر اول را از فایل jsonخوانده و با استفاده از کتابخانه hazmنرمال سازی،  
توکن سازی در سطح کلمات و ریشهیابی را انجام میدهیم. یک آرایه دوبعدی تعریف کرده که همان  
ماتریس رخداد کلمات و اسناد میباشد. بنابراین تعداد سطر های این ماتریس برابر با تعداد توکن های  
استخراج شده و تعداد ستونها برابر با تعداد اسناد( )۱۰۰۰میباشد شماره سطر نشان دهنده اندیس  
توکن در posting listو شماره ستون همان docIdاست(.در ابتدا تمام خانههای این ماتریس را برابر  
صفر قرار میدهیم) در ادامه با استفاده از خروجی تمرین قبل که همان posting listاست برای هر  
توکن اسناد مربوطه را پیدا کرده(اسناد مربوطه یعنی اسنادی که شامل آن توکن هستند) و درایه مربوط  
به توکن و سند مورد نظر را یک میکنیم.  
برای اینکه بدانیم هر سطر ماتریس مربوط به کدام توکن میباشد یک دیکشنری تعریف میکنیم و هر  
توکن را به اندیس مربوط به آن در posting listنگاشت میکنیم  
نحوه پاسخ به کوئری ها به شکل زیر است:  
برای پیادهسازی عملوند andسطرهای مربوط به توکنهای موجود در عملوند های کوئری را از ماتریس  
خوانده و در هم ضرب میکنیم. میدانیم که این ضرب، ضرب آرایه ای است نه ماتریسی یعنی تنها درایه  
های متناظر در هر بردار در هم ضرب شده و بردار نتیجه را میسازد(.در جبر بولی andمعدل ضرب  
است (برای پیادهسازی عملوند notسطر مربوط به توکن عملوند را خوانده و یک را از تمام درایه های آن کم  
میکنیم(.میدانیم در جبر بولی عمل notمعادل تفاضل یک از عملوند notاست)