ابتدا متن 100 خبر اول را در یک لیست ذخیره کردیم. به کمک کتابخانه Hazm، نرمالسازی و توکن سازی کردیم. توکن سازی یک لیست خروجی می‌دهد که شامل 100 عضو است و ایتم اول آن بعنوان مثال، شامل یک لیست از توکن‌های خبر اول است. تعداد کل توکن‌ها در همه خبرها (شامل تکراری‌ها) 43034 می‌باشد.

مرحله بعد ریشه یابی است. پس از ریشه یابی، با توکن های ریشه یابی شده‌ای که بیشتر از 3 حرف دارند، پستینگ لیست را می‌سازیم. در اخر تعداد توکن‌ها(غیر تکراری و بیشتر از 3 حرف) برابر 5277 شد.

پستینگ لیست ما مقابل هر توکن، داکیومنت آیدی‌هایی که در آن ها موجود است و تعداد تکرار آن در هر داکیومنت را شامل می‌شود. سپس تعداد تکرار توکن در کل داک‌ها را به لیست اضافه کردیم.