

بازیابی اطلاعات

گروه 2

محمد صادق سیرجانی

سید امیر موسوی

جاوید چاجی

شهریار بلورساز

هدف از فاز اول پروژه پیاده سازی پیش پردازش ها (جداسازی جملات و کلمات ،حذف ایست واژه ها, نرمال سازی و ریشه یابی و ریشه یابی معنایی) با استفاده از ابزار های پردازش زبان فارسی بود.

همچنین در ادامه به استخراج تعداد کلمات در اسناد پرداخته و پس از آن `posting_list` را برای ایجاد `index` های نهایی ایجاد کردیم.

پس از انجام موارد فوق دیتا `posting_list` ایجاد شده در غالب یک فایل ایجاد شده است.