

به نام خدا

گزارش کار تمرین درس بازیابی پیشرفته اطلاعات

محمد رازقی رستمی

شماره دانشجویی : ۹۰۱۰۶۰۲۸

در این تمرین ما یک موتور جستجو را با استفاده از فریم‌ورک Map-Reduce و ابزار hadoop پیاده سازی کردیم.

به طور کلی تمام کدهای نوشته شده گرفته از : ایندکس کردن، آپدیت کردن ایندکس، ساختن بردار طول برای هر داکيومنت ، جستجو به چهار روش ذکر شده در صورت پروژه، با استفاده از Map-Reduce انجام شده.

با اجرای فایل ui.sh شما می‌توانید واسط کاربری را مشاهده کنید که از شما به نوع جستجو و یک Query میگیرد و در نتیجه به شما داکيومنت‌های مربوط را نمایش می‌دهد. همچنان کار correction و پیشنهاد بهتر دادن را نیز انجام میدهد.

با اجرای فایل complete_index.sh فایل ایندکس ساخته شده و فشرده سازی می‌شود. برای فشرده سازی از روش dictionary_as_string استفاده شده است. که در زیر اندازه‌ی فایل قبل و بعد از فشرده‌سازی را می‌بینید :

995712 after

1007776 before

که به میزان ۹۸ درصد فشرده‌سازی انجام شده(البته با زیاد شدن تعداد اسناد این عدد کوچکتر می‌شود).

با استفاده از فایل evaluate_all.sh جستجو روی پرسش‌های نمونه انجام می‌شود و در جایی ذخیره می‌شود. سپس به وسیله‌ی فایل judge.sh می‌تواند درصد percision و recall را ببینید.

با تشکر فراوان