

گزارش فاز دوم پژوهشی بازیابی پیشرفته اطلاعات

سحر موالی

غزل ابراهیمی

بخش اول

در این بخش داده‌ها را از فایل‌ها خوانده و پیش‌پردازش‌های خواسته شده را روی آن‌ها انجام دادیم. برای این کار از کتابخانه‌های `hazm` و `nltk` استفاده کردیم.

بخش دوم

در این بخش توابعی برای ساختن نمایه‌ی `positional` و `bigram` پیاده‌سازی کردیم. این نمایه‌ها در فایل‌هایی با همین نام‌ها در کنار فایل‌های پژوهه ذخیره شده‌اند.

بخش سوم

در این بخش نمایه‌های ساخته شده در بخش قبل را با روش `gamma code` و `variable byte code` فشرده کردیم. نمایه‌های فشرده شده حاصل از اجرای این الگوریتم‌ها در فایل‌هایی با همین نام‌ها در کنار سایر فایل‌های پژوهه ذخیره شده‌اند.

بخش چهارم

در این بخش با استفاده از `jaccard` و `edit distance` معیار برای هر پرسمن و رودی حالت اصلاح شده را محاسبه می‌کنیم.

بخش پنجم

فقط الگوریتم اول را پیاده‌سازی کردیم. به این صورت که توابعی برای محاسبه‌ی `tf-idf` اسناد محاسبه کردیم. تابع `compute_cosine` با استفاده از این توابع و با گرفتن پرسمن به عنوان ورودی لیستی مرتب از اسناد را برمی‌گرداند.

مشارکت اعضای گروه

در انجام این فاز پژوهه اعضای گروه به صورت برابر همکاری کرده‌اند.