

## گزارش فاز دوم پروژه‌ی بازیابی پیشرفته‌ی اطلاعات

سحر موالی

غزل ابراهیمی

### بخش اول

در این بخش داده‌ها را از فایل‌ها خوانده و پیش‌پردازش‌های خواسته‌شده 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 پروژه ذخیره شده‌اند.

### بخش سوم

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

### بخش چهارم

در این بخش با استفاده از `edit distance` و معیار `jaccard` برای هر پرسمان ورودی حالت اصلاح‌شده 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 پرسمان به عنوان ورودی لیستی مرتب از اسناد را برمی‌گرداند.

### مشارکت اعضای گروه

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