هدف پیاده سازی یک موتور جستجو شامل بخشهای نمایه سازی، جستجو و رتبه بندی میباشد. ورودی این موتور حدود چهار هزار صفحه HTML به همراه آدرس میباشد که در تکلیف یک خزش شده است.

سناریوی کار به شکل زیر است:

الف) مرحله آفلاین:

ابتدا HTMLصفحات و آدرس آنها را از فایل خوانده و عملیات پارس را انجام میدهید. عملیات پارس شامل خواندن عنوان و تمام متن داخل صفحه (اعم از فارسی و انگلیسی) میباشد. سپس مرحله نمایه سازی انجام میگیرد (فقط یکبار انجام میشود و برای از بین نرفتن در فایل ذخیره میشود). نمایه به شکل معکوس ساخته میشود و شامل تمام متنهای همه بخشهای سند مانند عنوان و بدنه میباشد. همچنین لازم است علاوه بر نمایه سازی آدرس، عنوان و بدنه نیز ذخیره شوند. هنگام نمایه سازی لازم است نرمال سازی روی ی و ک حتما انجام شود و کارکترهای اضافه مانند ,;.*+-)(و ... حذف شود.

ب) مرحله آنلاین:

واسط کاربری تحت وب فارسی برای این موتور طراحی نمایید که شامل باکس جستجو و نمایش نتایج میباشد. شامل مراحل زیر میباشد:

- **کاربر پرس وجوی خود را وارد میکند (AND). (دانشگاه یزد)**
- **نتایج بعد از رتبه بندی شامل عنوان(قابل کلیک)، خلاصه ای از بدنه به کاربر نشان داده میشود.**
- **تعداد نتایج و زمان جستجو نیز نشان داده میشود.**
- **وزن دهی مناسب برای عنوان و بدنه**
- **خطایابی برای با یک خطا انجام شود و در صورت خطایابی عبارت جستجو شده نشان داده شود**
- **جستجوی wildcard پیشتیبانی شود**

توجه:

- **پیش فرض پرس وجو AND میباشد.**
- **هر ساختار و ساختمان داده ای را برای ذخیره و نگهداری در حافظه میتوانید استفاده نمایید.**
- **گزارشاتی مانند زمان و حجم نمایه سازی حتما قید شود.**
- **تحویل حضوری از دو مرحله تشکیل شده است: ارائه در قالب پاورپوینت (۵۰ درصد) و اجرای برنامه ۵۰ درصد**

پروژه به سه صورت میتواند پیاده سازی شود (مورد ۲ و ۳ پیشنهاد میشود):

۱) از پایه با استفاده از یکسری کتابخانه ها
۲) استفاده از Elastic Search
۳) استفاده از OpenSearch