

هدف پیاده سازی یک موتور جستجو شامل بخشهای نمایه سازی، جستجو و رتبه بندی میباشد. ورودی این موتور حدود ده هزار صفحه HTML به همراه آدرس میباشد که در قالب فایل XML تهیه شده است (به پیوست آمده است).

سناریوی کار به شکل زیر است:

الف) مرحله آفلاین:

ابتدا HTML صفحات و آدرس آنها را از فایل XML خوانده و عملیات پارس را انجام میدهند. عملیات پارس شامل خواندن عنوان و تمام متن داخل صفحه (اعم از فارسی و انگلیسی) میباشد. سپس مرحله نمایه سازی انجام میگردد (فقط یکبار انجام میشود و برای از بین رفتن 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.

ب) مرحله آنلاین:

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

- کاربر پرس وجوی خود را وارد میکند (AND یا OR). (دانشگاه یزد = دانشگاه AND یزد، دانشگاه OR یزد)
- نتایج بعد از رتبه بندی با فرمول TFIDF شامل عنوان (قابل کلیک)، خلاصه ای از بدنه به کاربر نشان داده میشود.
- تعداد نتایج و زمان جستجو نیز نشان داده میشود.
- همچنین صفحه بندی نیز وجود دارد.
- وزن دهی مناسب برای عنوان و بدنه

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 نمایه سازی حتما قید شود.
- برای سادگی در نمایه سازی از روش non-positional استفاده گردد.
- تحویل حضوری از دو مرحله تشکیل شده است: ارائه در قالب پاورپوینت و اجرای برنامه ۵۰ درصد، اجرای برنامه ۵۰ درصد

