

باسمه تعالی

پروژه دوم درس بازیابی اطلاعات
مدرس: سعید شهریوری

در این پروژه شما یک موتور جستجوی ساده تولید خواهید کرد. رویکرد این موتور جستجو پوشش حداکثر ۱۰۰ هزار صفحه فارسی است. مواردی که باید در این موتور جستجو در پیاده‌سازی شوند از قرار زیر است:

- یک خزشگر ساده برای دریافت صفحات وب و یک پارسر ساده برای استخراج متون از صفحات و شروع از صفحه اصلی دانشگاه ارومیه به عنوان هسته اولیه
- یک مخزن اسناد مبتنی بر پایگاه داده برای ذخیره‌سازی اسناد خزش شده
- استفاده از الگوریتم‌های مناسب برای رتبه‌بندی نتایج (برای مثال tf-idf)
- طراحی یک واسط کاربری وب ساده
- ارائه واسط‌های مناسب Restful برای جستجو، درج، ویرایش، و حذف صفحات خزش شده
- امکان تصحیح پرس‌وجوهای کاربر

موارد زیر دارای نمره اضافی می‌باشند:

- استفاده از Elasticsearch برای پیاده‌سازی این موتور جستجو
- پیاده‌سازی سیستم Caching برای پرس‌وجوها
- امکان جستجوی تصویر
- تحویل پروژه در قالب یک تصویر داکر (Docker Image)

لطفاً به نکات زیر توجه کنید:

- آخرین مهلت تحویل پروژه ساعت ۲۴ روز ۳۰م دی ماه است و در صورت عدم تحویل در موعد مقرر نمره‌ای به شما تعلق نمی‌گیرد.
 - انجام پروژه در قالب گروه‌های ۳ یا ۴ نفره آزاد است.
 - در صورتی که درس مهندسی اینترنت را نیز اخذ کرده‌اید با انجام این پروژه می‌توانید از نمره آن به عنوان پروژه دوم آن درس نیز استفاده کنید. البته برای این منظور سیستم شما باید در زمینه پیاده‌سازی واسط‌های REST و Caching کامل باشد.
- موفق باشید