

پروژه دوم درس بازیابی پیشرفته اطلاعات

پیاده‌سازی یک متاموتور جستجو

موتورهای جستجوی موجود در وب را می‌توان به دو دسته کلی تقسیم کرد. موتورهای مبتنی بر پیمایشگر (crawler based search engines) و متاموتورهای جستجو (meta search engine). دسته اول دارای یک ماجول بنام پیمایشگر (یا Spider) هستند که کار کشف، جمع‌آوری و شاخص‌گذاری صفحات وب را انجام می‌دهد در حالی که دسته دوم فاقد این بخش هستند. نحوه عمل متاموتورها به این صورت است که درخواست کاربر را به تعدادی از موتورهای مبتنی بر پیمایشگر می‌فرستند و پس از دریافت نتایج آنها، با استفاده از یک الگوریتم خاص نتایج مختلف را رتبه‌بندی کرده و به عنوان خروجی به کاربر نشان می‌دهند.

هدف از این پروژه پیاده‌سازی یک متاموتور جستجو می‌باشد. در حال حاضر تعداد زیادی موتور جستجوی تجاری در وب وجود دارد که اغلب آنها امکان ارسال درخواست کاربر و دریافت نتایج را بوسیله API های مشخص فراهم کرده‌اند. نحوه انجام این پروژه به صورت گروهی است. اعضای هر گروه در وهله اول باید ۳ موتور را به عنوان موتورهای پایه‌ای خود انتخاب کنند. سپس یک الگوریتم برای تلفیق نتایج دریافت شده از موتورهای سه‌گانه پیشنهاد نموده و پیاده نمایند (از هر موتور ۱۰ نتیجه اول دریافت می‌شود). در مرحله آخر نیز باید یک واسط کاربر مناسب با امکانات لازم (امکان استفاده از عملگرهای منطقی، +، - و ") تهیه شود.

نکات قابل توجه :

- تعداد اعضای مجاز در هر گروه حداکثر دوف نفر می‌باشد.
- لیستی از موتورهای جستجوی قابل استفاده را می‌توانید در آدرس زیر پیدا کنید:
- <http://www.search-engines-megalist.com/>
- به علت کیفیت متفاوت موتور google هیچ کدام از گروه‌ها نباید از گوگل استفاده نماید.
- تحویل این پروژه به صورت حضوری می‌باشد.

- قبل از تحویل حضوری پروژه هر گروه باید در قالب یک گزارش ۳ صفحه‌ایی، به صورت خلاصه در مورد موتورهای انتخاب شده، الگوریتم پیشنهادی و تحلیل نتایج (مقایسه با نتایج موتورهای مورد استفاده) توضیحاتی ارائه کند.
- بخش عمده‌ایی از نمره این پروژه به کیفیت الگوریتم تلفیق و امکانات واسط کاربر اختصاص دارد.
- موعد نهایی تحویل این پروژه ۸۴/۹/۷ می باشد. این مهلت تحت هیچ شرایطی تمدید نخواهد شد.
- برای پرسیدن سوالات خود می‌توانید از بخش Discussion Area در سایت درس استفاده نمایید.