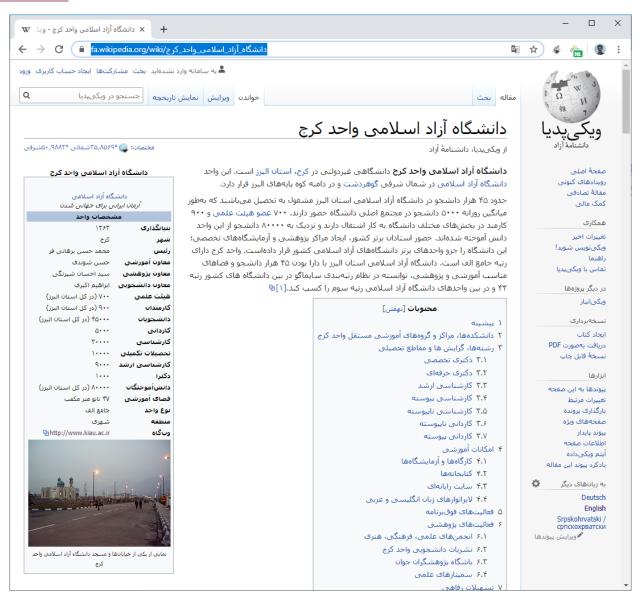
پروژه پایانی درس «مبانی بازیابی اطلاعات و جستجوی وب» نیمسال ۱-۹۹-۹۸

هدف: طراحی یک سیستم کوچک جستجوی اطلاعات راجع به دانشگاه آزاد اسلامی واحد کرج از سایت ویکیپدیا

صفحه دانشگاه آزاد اسلامی واحد کرج در سایت ویکیپدیا:

https://fa.wikipedia.org/wiki/%D8%AF%D8%A7%D9%86%D8%B4%DA%AF%D8%A7%D9%87_%D8%A2%D8%B2%D8 %A7%D8%AF_%D8%A7%D8%B3%D9%84%D8%A7%D9%85%DB%8C_%D9%88%D8%A7%D8%AD%D8%AF_%DA%A9%D8%B1%D8%AC



شرح پروژه:

با استفاده از «ابزار هضم» ام مجموعه دستورات و برنامههای لازم، سیستم بازیابی اطلاعاتی طراحی کنید که توکنها (نشانههای) موجود در یک صفحه فوق که صفحه «دانشگاه آزاد اسلامی واحد کرج» در ویکیپدیا است را استخراج کرده، پس از انجام عملیات ریشه گیری کلمات فارسی، یک ساختار شاخص موقعیتی (Positional Index) بسازد.

این سیستم باید امکان جستجو با عملگرهای AND و OR و جستجوی شامل عبارات دقیقی مانند "مهندسی کامپیوتر" و "نشریات دانشگاهی" در این شاخص موقعیتی را داشته باشد. برای سادگی، در هر عبارت پرسوجو، حداکثر یک عملگر AND یا OR استفاده می شود و اگر عملگر "" (Double Quotation) استفاده شود، عملگرهای AND و OR در پرسوجو مجاز نیستند.

نشان دهید که آیا عبارات زیر در این سند وجود دارند یا خیر:

نتيجه	عبارت پرسوجو
بله (به همراه تو کنهای چپ و راست)	"مهندسي كامپيوتر"
بله (به همراه تو کنهای چپ و راست)	"فیش هفتگی"
خير	هیچ AND دانشگاه
بله	حضور AND اقتصاد
بله	هیچ OR دانشگاه
بله (به همراه تو کنهای چپ و راست)	"مجاز به انتشار آن به"
بله (توضیح: در متن بصورت جمع آمده و طی ریشه گیری به مفرد تبدیل شده است)	سمينار

در صورتی که عبارت پرسوجو شامل عملگر "" (Double Quotation) باشد، در کنار نتیجه ی نمایش داده شده، یک تو کن از سمت چپ و یک تو کن از سمت راست از داخل ساختار شاخص (و نه از درون سند) باید استخراج شده و نمایش داده شود. برای مثال، در صورت جستجوی عبارت "فیش هفتگی"، علاوه بر اینکه نتیجه «بله» باید نمایش داده شود، تو کنهای «تهیه» و «از» نیز باید به شیوه مناسبی نمایش داده شود، زیرا عبارت «تهیه فیش هفتگی از» در متن این سند هست.

پرینت متن برنامه و هفت تصویر (Screenshot) از نتیجه اجرای برنامه خودتان برای هفت پرسوجوی جدول فوق را حداکثر تا پایان آذرماه ۱۳۹۸ تحویل دهید و همچنین متن برنامه و تصاویر را به آدرس hsaboohi@gmail.com ایمیل کنید. (توجه بفرمایید: پرینت و ایمیل، هر دو لازم هستند)

موفق باشید هادی صبوحی

¹ http://www.sobhe.ir/hazm/