قسمت اول

در این پروژه یک موتور جستجو برااییکی ازسایتهایمعروفزومیت(https://www.zoomit.ir/) یامجله

فرادرس(https://blog.faradars.org) ایجاد می کنیم. شما باید یک خزنده بنویسید که ازآدرس اصلی داده شدهآغاز به

کار کند و ارجاعاتی که در همین دامنه هستند را دنبال کند و حداقل100صفحه را جمع آوری کند.

در سایت زومیت درمنوی فناوری، هر بخش مسیر مشخصی دارد (مثالhttps://www.zoomit.ir/computer-

(/learningاکه اگر صفحاتی ازسه مورد از این زیر بخش ها را جمع کنید در انجام قسمت دوم تمرینکار شما تسهیل خواهد شد.

درسایت مجله فرادرس هر مقاله یک دستهبندی دارد که باالی عنوان به صورت لینک آبی نوشته شده است. در اینجا نیز اگر

مقاالتی از سه دسته را جمع کنید در انجام قسمت دوم تمرین کار شما تسهیل خواهد شد.

بخش های مختلف مفید هر صفحه مانند عناوین و متنو...را جمع آوری کنید. برای پردازش صفحاتHtmlمی توانید از کتابخانه های مناسب موجود استفاده کنیدو محتوای بخشهای مختلف صفحه را استخراج کنید.

سپس محتوای جمع شده را با استفاده از elastic search یا برنامه خودتان که درتمرین قبل با آن کار کرده اید شاخص گذاری کنید. با وارد نمودن تعدادی پرس و جوی دلخواه عملکرد مناسب جستجو را نشان دهید

.به صورت اختیاری بر اساس بخشهای مختلف صفحه مکانیزم رتبه بندی ایجاد کنید. به عنوانمثال اگر در سندی کلمه در عنوان متن ظاهر شده بود، این سند در رتبهباالترینسبت به سندی که کلمه را در متن خود داردقرار بگیرد. با وارد نمودن تعدادی پرس و جوی دلخواه عملکرد مناسب جستجو را نشان دهید.نحوه پیاده سازی خزشگر وموتور جستجو، به همراه انواع پرس و جوها که قابلیت برنامه شما را نشان میدهند را در گزارش نهایی ذکر کنید.

قسمت دوم

صفحات مرتبط با حداقل سه بخش مختلف سایت(که درقسمت اول شرح داده شد)را خزش کنید. از هر بخش حداقل 25 صفحه را خزش کنید.سپس برای صفحات هر بخش برچسب همان بخش را در نظر گرفته و یک رده بند بیز ساده را برای

رده بندی صفحات اجراکنید.به این منظور می توانید از کتابخانه های هر زبانی که استفاده میکنید بهره ببرید یا از ابزارهای آماده مثل وکا استفاده کنید.شرح کار و دقت رده بند را گزارش کنید.

به صورت اختیاری میتوانید صفحات را خوشه بندی کنید وخوشه های حاصل شده را با برچسب واقعی اسناد مقایسه کرده، معیارهای ارزیابی مناسب را محاسبه کنید.

و در آخر نیاز به یک ویدئو دارم که ضبط بکنید از صفر تا صد پروژه رو توضیح بدید
