

مبانی بازیابی اطلاعات و جست و جوی وب

تمرین شماره 3

در این تمرین شما باید از مفاهیمی که در مبحث خزنده ها یاد گرفته اید استفاده کنید تا اطلاعات خاصی را از یک سایت crawl کنید.

- برای crawl کردن می توانید از ابزار ها و کتاب خانه های مختلفی مثل Scrapy و Selenium استفاده کنید.
- برای این تمرین باید تمام course های موجود در [این سایت](#) را crawl کنید.
- ممکن است به دلایل مختلفی crawler شما نتواند همه course های سایت را crawl کند که یکی از وظایف شما در این تمرین بهبود دادن crawler خود به منظور crawl کردن حداکثر صفحات موجود است.
- اطلاعاتی که در این تمرین crawl می کنید برای تمرین بعدی شما که جست و جو کوئری های مختلفی بر روی داده های course ها است استفاده می شود. به همین خاطر سعی کنید که خروجی crawler شما شامل تمام course های سایت و اطلاعات مهم این course ها باشد تا در قسمت بعدی با مشکلی مواجه نشوید.

خروجی بدست آمده را در سایتی مانند dropbox یا google drive آپلود کنید و در آخر یک داکيومنت شامل قطعه کد مربوط به هر قسمت از برنامه، توضیح کد، نمونه ای از خروجی crawler خود و لینک فایل های آپلود شده را به همراه کد crawler در یک فایل zip ارسال کنید.