

عنوان کارگاه: خزش وب و شبکه‌های اجتماعی

ارائه دهندگان: رامین رضوانی، سجاد رضایی، سمیه سلطانی

چکیده: بالا رفتن قدرت محاسباتی سیستم‌ها، افزایش حافظه‌ها و همچنین دسترسی آسان به داده‌های موجود در وب، باعث شده است تا پژوهشگران حوزه علوم داده به استفاده از داده‌های حجیم، علاقه مند شوند. با توجه به این که سایت‌ها، شرکت‌ها و بنگاه‌های تجاری اینترنتی، داده‌های خود را به سادگی در اختیار دیگران نمی‌گذارند، پژوهشگران و برنامه نویسان، روش‌ها و ابزارهای قدرتمندی را برای جمع‌آوری و ذخیره‌سازی داده‌ها توسعه داده‌اند که به آن خزش یا کراال می‌گویند. با استفاده از این ابزارها، داده‌ها و اطلاعات داخل وب سایت‌ها و شبکه‌های اجتماعی را بدست آورده و در قالب‌های معمول و قابل استفاده برای تحلیل‌های خود ذخیره می‌کنند.

در این کارگاه قصد داریم به آموزش وب اسکرپینگ (Web Scraping) برای صفحات وب و خزش شبکه‌های اجتماعی بپردازیم. اسکرپیر در واقع یک برنامه کامپیوتری (در این آموزش به زبان پایتون) است که خیلی ساده یک درخواست را به یک سرور ارسال می‌کند و پاسخ سرور را تجزیه و بخش مورد نظر ما را چاپ می‌کند. در بخش خزش شبکه‌های اجتماعی نیز روی دو مورد اینستاگرام و تلگرام، به نحوه ساخت یک خزشگر به زبان پایتون خواهیم پرداخت.

سرفصل مطالب:

✓ خزشگر وب

۱. معرفی web scraping و کاربرد آن
۲. معرفی کتابخانه requests در پایتون
۳. نحوه ارسال درخواست HTTP و دریافت پاسخ آن
۴. معرفی ابزار selenium در پایتون و کاربرد آن در خزش صفحات وب
۵. معرفی کتابخانه BeautifulSoup
۶. استفاده از کتابخانه BeautifulSoup برای خواندن تگ HTML از صفحه وب و نمایش آن
۷. نمونه عملی خزشگر وب

✓ خزشگر اینستاگرام:

۱. معرفی اصطلاحات مهم و پایه در اینستاگرام
۲. اطلاعات قابل جمع‌آوری در اینستاگرام
۳. روش‌های جمع‌آوری اطلاعات از اینستاگرام
۴. نمونه عملی از خزش اینستاگرام

✓ خزشگر تلگرام:

۱. مقدمات خزش شبکه‌های اجتماعی
۲. آشنایی با خزش تلگرام
۳. معرفی کتابخانه Telethon
۴. نحوه ساخت ربات‌های تلگرام