

خزشگر اینستاگرام

در این پروژه شما باید با استفاده از کتابخانه های زیر یک خزشگر اینستاگرام بنویسید. :

- BeautifulSoup
- requests
- Selenium
- Tkinter
- pandas
- threading

استفاده از بسته های دیگر در این پروژه مجاز نمی باشد.

برنامه شما باید حاوی بخشهای زیر باشد:

۱. یک هشتگ دلخواه را در اینستاگرام جست و جو کند و n اکانتی را که در نتایج جست و جوی اینستاگرام حاوی این هشتگها بوده اند لیست کند. n باید پارامتریک باشد و ابتدای برنامه قابل تنظیم باشد.

۲. در مرحله بعد m پست آخر هر یک از این اکانتها را در نظر بگیرید و متن کامنتهای ذیل هر کدام از این پست ها را به همراه تعداد لایک های آن استخراج کنید. m باید به صورت پارامتری قابل تنظیم باشد.

۳. اطلاعات ذخیره شده را در یک دیتافریم و نهایتا روی هارد ذخیره کنید. دیتافریم شما باید حاوی کامنت، نام کاربری نویسنده کامنت T نام کاربری اکانت پست اصلی و تعداد لایکهای کامنت باشد.

۴. یک واسط کاربری گرافیکی (که ترجیحا با `tkinter` ایجاد شده باشد) که دارای ابزارهای لازم برای تعامل با کاربر و نمایش خروجی های مورد نظر باشد. از جمله:

- در این پنجره کاربر باید بتواند لیست اکانتهایی را که برنامه شما سراغ پست های آنها خواهد رفت را ببیند و آنها را کم یا زیاد کند.
- مقدار m و n را تنظیم کند.
- همچنین باید در این پنجره به کاربر نشان داده شود که تا کنون چه تعداد از اکانتها خزش شده اند و چه تعداد باقی مانده است.
- زمان سپری شده و زمان تخمینی تا انتهای کار نیز باید نمایش داده شود.
- پس از پایان کار نیز گزارشی از زمان سپری شده برای کل کار را نشان دهد.
- محل ذخیره فایل خروجی روی هارد توسط کاربر تعیین شود.
- اضافه کردن موارد دیگر در این واسط گرافیکی نمره امتیازی خواهد داشت.

۵. برنامه شما باید به صورت مالتی ترد نوشته شود. می توانید انتخاب کنید که برای خزش هر اکانت از یک ترد استفاده کنید یا برای خزش هر کامنت یک ترد جدید ایجاد کنید. در صورتی که اجرای برنامه شما از سایر هم کلاسی هایتان سریع تر باشد، نمره امتیازی خواهد داشت.

۶. در این برنامه استایل برنامه نویسی شی گرا مد نظر نیست؛ اما در صورت پیاده سازی این برنامه به صورتی شی گرا، نمره امتیازی خواهد داشت.

۷. توابع و کلاسهایی که تعریف میکنید باید دارای داک استرینگ باشند. تمیز بودن کدها طبق اصول معرفی شده در کلاس درس الزامی است.

بخش امتیازی ویژه

یک مساله خاص را در نظر بگیرید و هشتگ های مرتبط با آن را از طریق برنامه خودتان جست و جو کنید. مثلاً فرض کنید می خواهید ببینید کامنتهای افراد در مورد شرکت سامسونگ چه قدر مثبت یا منفی است. تمام هشتگ های مربوط به شرکت سامسونگ از جمله انواع برندهای مربوطه و ... را از طریق برنامه خودتان جست و جو کنید کامنتهای مربوطه را استخراج کنید.

سپس تعداد 1000 تا از کامنتها را به صورت تصادفی درنظر بگیرید و آنها را برچسب گذاری کنید. به این صورت که اگر کامنت دارای نظر مثبت نسبت به شرکت سامسونگ بود، برچسب مثبت، در صورتی که دارای نظر منفی بود، برچسب منفی و در غیر این صورت دارای برچسب خنثی باشد. (به صورت معمول برچسب گذاری 1000 تا کامنت کمتر از 2 ساعت از شما زمان می گیرد. می توانید این بخش از کار را به کمک سایر همکلاسی هایتان انجام دهید. هر چه تعداد کامنتها در این بخش بیشتر باشد، دقت خروجی شما بیشتر می شود)

با استفاده از ماژول `fasttext` در پایتون می توانید یک مدل بسازید که از روی این 1000 کامنت برچسب خورده تا حدی الگوی نظرات مثبت و منفی را یاد بگیرد. سپس این مدل می تواند با درکی که نسبت به منفی یا مثبت بودن یک نظر پیدا کرده، نظر خودش را درباره مثبت و منفی بودن هر کامنت جدیدی اعلام کند! بنابراین می توانید با این مدل تمام نظرات را تست کنید و بررسی کنید چه میزان از نظرات مثبت یا منفی بوده اند.

برنامه شما میتواند گزارش کند که چه تعداد از نظرات کاربران درباره این موضوع مثبت یا منفی بوده است. (یا به صورت درصد نمایش دهد)

توضیحات لازم برای انجام دادن این بخش از پروژه را می توانید در [این صفحه](#) ببینید.

گزارش نویسی

لازم است یک فایل گزارش در مورد پروژه خود، ویژگی ها و بخشهای مختلف آن نیز آماده کنید. فایل گزارش، به همراه کدهای پروژه، داده های استخراج شده و نتایج یک نمونه از اجرای برنامه را در قالب یک فایل zip اپلود کنید. استفاده از Git دارای نمره امتیازی است.

زمان ارسال فایل نهایی

شما تا ساعت 23:59 چهارم بهمن ماه 1400 فرصت دارید، فایل نهایی خود را بارگذاری کنید. پنجم بهمن ماه ارائه شفاهی پروژه شما خواهد بود که درباره زمان بندی آن بعداً هماهنگ خواهیم کرد.

با آرزوی موفقیت