

هدف این تمرین آشنایی و تعامل با ابزارها و خزنده‌های^۱ شبکه‌های اجتماعی است.

۱) کار با ابزار گفی^۲

با استفاده از نرم‌افزار گفی توئیت‌های شبکه‌ی اجتماعی توئیتر با یک هشتگ دلخواه را جمع‌آوری و شبکه‌ی به‌دست آمده را ترسیم کنید. مراحل انجام کار را توضیح دهید.

a. نمودار توزیع درجات شبکه را ترسیم کنید.

b. معیارهای مرکزیت بینایی^۳، مرکزیت نزدیکی^۴، مرکزیت هارمونیک^۵ و پیچ‌رنک را محاسبه و گزارش کنید.

c. انجمن‌های شبکه‌ی حاصل را مشخص کنید و با چیدمان^۶ مناسب این انجمن‌ها را نشان دهید.

۲) توئیتر

ا) مراحل استفاده از API توئیتر را توضیح دهید. سپس کدی بنویسید که لیستی از شناسه‌ی توئیت‌ها را دریافت کند و اطلاعات مربوط به آن توئیت را جمع‌آوری و ذخیره کند. همچنین در صورتی که توئیت حاوی عکس باشد آن عکس را ذخیره کند.

ورودی: فایلی به نام tweets_id.csv است که هر سطر آن شامل یک شناسه توئیتری مرتبط با یک توئیت در شبکه‌ی اجتماعی توئیتر است.

خروجی: ۱) چند فایل متنی به تعداد شناسه‌های موجود در فایل ورودی، که نام آن‌ها همان شناسه‌ی توئیت است. محتوای هر فایل نیز اطلاعات توئیت مربوطه می‌باشد که در قالب json ذخیره شده است. ۲) اگر توئیتی دارای عکس باشد، آن عکس را با نام شناسه‌ی توئیت ذخیره کنید. ۳) یک لیست از شناسه‌ی کاربران منتشرکننده‌ی این توئیت‌ها (این لیست به عنوان ورودی بخش «ب» و «پ» مورد استفاده قرار خواهد گرفت).

تذکر: یکی از شناسه‌های درج‌شده در فایل ورودی، مربوط به توئیتی غیرقابل دسترسی است که به دریافت خطا منجر خواهد شد. نوع این خطا را گزارش کنید و بروز این خطا را به‌گونه‌ای مدیریت کنید که خللی در روند جمع‌آوری بقیه‌ی توئیت‌ها ایجاد نشده و برنامه متوقف نشود.

¹ Crawler

² Gephi

³ Betweenness Centrality

⁴ Closeness Centrality

⁵ Harmonic Centrality

⁶ Community

⁷ Layouts

ب) کدی بنویسید که لیستی از شناسه‌های کاربری بگیرد (خروجی ۳ از بخش «آ» همین سوال) و اطلاعات پروفایل کاربران را جمع‌آوری کند.

خروجی: چند فایل متنی به تعداد شناسه‌های موجود در لیست ورودی، که نام هر فایل همان شناسه توئیتر کاربر است. محتوای هر فایل نیز اطلاعات پروفایل کاربر مربوطه می‌باشد که در قالب json ذخیره شده است.

پ) کراولری بنویسید که لیستی از شناسه‌های کاربری بگیرد (خروجی ۳ از بخش «آ» همین سوال) و اطلاعات شبکه‌ای کاربران را جمع‌آوری کند. (بطور مثال لیستی از نام دوستان و دنبال‌کننده‌های کاربر)

نکته: برای این کار کفایت تنها ۲۰ دنبال‌کننده^۸ و دوستان^۹ کاربر مرتبط را استخراج کنید.

خروجی: چند فایل به تعداد شناسه‌های موجود در لیست ورودی، که شامل اطلاعات شبکه‌ای کاربران می‌باشد (یک فایل برای دوستان و یک فایل برای دنبال‌کننده‌ها). نام فایل‌ها نیز شامل همان شناسه توئیتر کاربر است.

ت) در این بخش می‌بایست بصورت بلادرنگ توئیتهای را بر اساس کلمات کلیدی و موقعیت مکانی فیلتر کنید. تفاوت نیازمندی‌های این بخش را نسبت به ۲ بخش قبل تبیین کنید.

- توئیتهایی که شامل یکی از دو کلمه کلیدی vaccination یا tokyo2020 هستند را جمع‌آوری کنید.
- توئیتهایی را جمع‌آوری کنید که در یک ناحیه جغرافیایی داخل ایران صادر شده‌اند. مختصات جغرافیایی این منطقه بصورت نقاط یک مستطیل است.

مختصات اول رأس چپ-بالا (47.526684 , 36.385913) و مختصات دوم رأس راست-پایین (58.512660 , 30.524413) مستطیل را مشخص می‌کند. نقاط گفته‌شده در تصویر زیر مشخص شده‌اند:



ورودی: کلمات کلیدی و مختصات ذکر شده در تمرین.

خروجی: دو فایل، که فایل اول شامل حداقل ۵۰ توئیتهای فیلترشده بر اساس کلمات، و فایل دوم نیز شامل حداقل ۵۰ توئیتهای که بر اساس مختصات جغرافیایی فیلتر شده‌اند.

نکته: فایل اول را `filtered_by_keywords.txt` و فایل دوم را `filtered_by_geo.txt` نامگذاری کنید.

ث) کراولری بنویسید که به تعداد ۲۰ توئیتهای جمع‌آوری کند که در متن این توئیتهای کلمه‌ی `Iran` به کار رفته باشد (نه الزاماً به عنوان هشتگ).

خروجی: فایلی با فرمت json شامل اطلاعات ۵۰ توئیتهای خواسته‌شده و با نام `iran.json`.

راهنمایی: می‌توانید از کتابخانه [Tweepy](#) در زبان پایتون استفاده کنید و از `python jupyter notebook` جهت تست و بررسی ساده‌تر استفاده کنید.

تذکر: دریافت کلید API توئیتر زمان‌بر است لذا انجام این تمرین را به تعویق نیندازید.

⁸ Followers

⁹ Friends

۳) اینستاگرام

آ) برنامه‌ای بنویسید که لیست شناسه کاربران را از داخل یک فایل بخواند و اطلاعات زیر را جمع‌آوری کند:

- **اطلاعات پروفایل کاربران** (عکس پروفایل، نام، تعداد پست‌ها، تعداد دنبال‌کننده‌ها، تعداد دنبال‌شونده‌ها و غیره)
- **اطلاعات پست‌های کاربران** (تنها پست‌های قابل مشاهده در صفحه اینستاگرام کاربر کفایت)
- **نظرات آخرین پست هر کاربر** (اطلاعات حداقل ۱۰ نظر از آخرین پست منتشرشده توسط هر یک از کاربران، در صورت قابل دسترسی بودن، را جمع‌آوری و ذخیره کند. اگر نظرات آخرین پست غیرقابل دسترسی بودند، نظرات یکی از پست‌های قبل از آن را ذخیره کنید)

ورودی: یک فایل به نام insids.csv که هر خط آن شامل یک شناسه کاربری می‌باشد.

خروجی: خروجی بخش‌ها شامل چند پوشه می‌باشد که نام هر پوشه، شناسه اینستاگرامی کاربر مربوطه است. هر پوشه متشکل از یک فایل شامل اطلاعات پروفایل کاربری، عکس پروفایل کاربر، یک فایل شامل اطلاعات پست‌های کاربر، یک پوشه شامل عکس‌های پست‌های او و یک فایل شامل اطلاعات حداقل ۱۰ نظر از آخرین پست آن کاربر است.

ب) برنامه‌ای بنویسید که لیست هشتک‌های اینستاگرام را از داخل یک فایل بخواند و **پست‌های قابل مشاهده در صفحه** حاصل از لیست پست‌های مرتبط با آن هشتک را جمع‌آوری کند.

ورودی: یک فایل به نام inshashtags.csv که هر خط آن شامل یک هشتک اینستاگرام است.

خروجی: متشکل از یک پوشه به ازای هر هشتک است که این پوشه شامل چند پوشه می‌باشد. هر پوشه بیانگر اطلاعات پست‌های مربوط به آن هشتک است. داخل این پوشه‌ها یک فایل شامل محتوای پست و عکس یا ویدیوی مربوط به آن پست می‌باشد. نام پوشه‌ها شناسه پست مربوطه است. به عنوان نمونه آدرس یک فایل شامل محتوای یک پست به شکل زیر می‌باشد:

شکل آدرس فایل اطلاعات پست:

hashtag_results/_hashtag_line_number/_post_id/_post_id_.txt

به عنوان مثال:

hashtag_results/1/234567/234567.txt

شکل آدرس عکس/ویدئو مرتبط با پست:

hashtag_results/_hashtag_line_number/_post_id/_post_id.png
hashtag_results/_hashtag_line_number/_post_id/_post_id.mp4

به عنوان مثال:

hashtag_results/1/234567/234567.png

راهنمایی: می‌توانید از کتابخانه InstagramScrapper در زبان پایتون استفاده کنید.

