



درس تحلیل شبکه‌های اطلاعات

مدرس: دکتر مصطفی صالحی

نیم سال اول ۹۹-۱۳۹۸

تمرین سری دوم

مهلت ارسال تمرین: ۱۳۹۸/۰۸/۱۴ تا

ساعت ۲۳

پایگاه داده‌های تحلیل اطلاعات

تحويل حضوری: اعلام خواهد شد

نکته: از هر منبعی که برای انجام این تمرین‌ها استفاده می‌کنید حتماً در گزارش خود ذکر کنید.

(۱) هدف از این تمرین، آشنایی با پایگاه داده‌ی ElasticSearch است

پایگاه داده‌ی ElasticSearch را روی سیستم خود نصب و راه‌اندازی کنید.

برای این منظور می‌بایست علاوه بر فایل گزارش یک فایل ReadMe که نحوه نصب و راه‌اندازی این پایگاه داده را مشخص می‌کند، آماده کنید.

توجه: در صورت استفاده از هرگونه منبع خارجی (سایت یا مقاله یا کتاب) حتماً در فایل گزارش خود ارجاع دهید.

راهنمایی: برای دریافت اطلاعات بیشتر به سایت زیر مراجعه فرمایید.

[TowardsDataScience : ElasticSearch with Python](#)

توجه: تمرین دوم و سوم به هم وابسته هستند. لطفاً هر دو تمرین را مطالعه و سپس شروع به انجام آنها کنید.

(۲) هدف از این تمرین، آشنایی با نوشتن کراولر برای شبکه‌ی تلگرام است.

پیش‌نیاز تمرین: آشنایی زبان برنامه نویسی پایتون

الف) به سایت تلگرام ([احراز هویت تلگرام](#)) و شماره خود را رجیستر کرده و مقادیر `api_id` و `api_hash` دریافت کنید.

ب) با استفاده از کتابخانه‌های موجود تلگرام^۱ (پیشنهاد می‌شود از تلتون^۲ استفاده کنید) برای شماره خود برنامه‌ای بنویسید و پست‌های دریافتی از شبکه‌ی تلگرام را کراول کرده و اطلاعات کاربران و پست‌های مربوط به آنها را استخراج کند.

^۱ api

^۲ Telethon (<https://github.com/LonamiWebs/Telethon>)

ج) سعی کنید برنامه بصورت بلادرنگ³ بنویسید که با دریافت هر پست جدید اطلاعات مربوط به پست را استخراج و به پایگاه داده ارسال نماید.

راهنمایی: برای دریافت اطلاعات بیشتر به سایت زیر مراجعه فرمایید.

[Telethon github Telethon Docs](#)

توجه: در صورت استفاده از هرگونه منبع خارجی (سایت یا مقاله یا کتاب) حتما در فایل گزارش خود ارجاع دهید.

۳) هدف از این تمرین، آشنایی با ذخیره‌سازی اطلاعات در پایگاه داده‌ی Elasticsearch است.

اطلاعات استخراج شده در تمرین دوم را در پایگاه داده‌ای که در تمرین اول راه‌اندازی کرده‌اید، ذخیره کنید. در فایل گزارش خود نحوه برقراری ارتباط با پایگاه داده و ذخیره‌سازی داده‌ها در آن را بنویسید.

۴) [اختیاری]: از پایگاه داده شامل خروجی‌های تمرین دوم و سوم، Export تهیه کنید و آن را کنار فایل گزارش خود قرار دهید. نحوه انجام این بخش را در گزارش خود ذکر کنید.

³ stream

گزارش کار:

گزارش کارهای خود را در قالب یک فایل PDF بنویسید. در صورتیکه از مرجعی برای نوشتن تمرین‌ها استفاده کرده‌اید، حتما در فایل گزارش به آن ارجاع دهید. فایل گزارش باید شامل نام، شماره دانشجویی و متن گزارش به تفکیک هر سؤال باشد. گزارش کار بخشی از نمره نهایی این تمرین خواهد بود. لذا در نگارش آن نهایت دقت را داشته باشید.

نحوه ارائه:

حتما به تمام بخش‌های تمرین خود مسلط باشید. در زمان ارائه ممکن است از شما توضیحاتی تکمیلی در مورد تمرین و یا تغییراتی در کد خواسته شود. نمره‌ی اختصاص داده شده برای افرادی که مسلط به کار خود باشند قابل تغییر است. در زمان مشخص شده جهت ارائه حتما حضور داشته باشید. غیبت در روز ارائه به منزله عدم ارسال تمرین است و نمره‌ای به آن تعلق نخواهد گرفت.

نکات تکمیلی:

در صورت تأخیر تا ۱۲ ساعت، ۲۵٪ و تا ۲۴ ساعت ۵۰٪ از نمره تمرین کسر خواهد شد. به تمرین با تأخیر بیشتر از ۱ روز نمره‌ای تعلق نخواهد گرفت.

آدرس ایمیل جهت ارسال تمرین: mostafa_salehi@ut.ac.ir و mh.qanbari@ut.ac.ir

فایل ارسالی یک فایل فشرده شده با قالب نامگذاری زیر باشد. (بجای NAME نام خود و به جای STDID شماره دانشجویی خود را قرار دهید).

hw2_NAME_STDID.zip

فایل فشرده شده باید شامل موارد زیر باشد:

۱- فایل‌های حاوی کدهای اجرایی (یک فایل ReadMe برای تمرین اول و چهارم)

۲- گزارش کار در قالب یک فایل PDF

موفق باشید. ☺