

تمرین میان ترم درس تحلیل اکتشافی داده ها با پایتون:

مهلت ارسال راه حل: نیمه شب 1401/01/24

برنامه ای (یک فانکشن با نام process) بنویسید که با دریافت آدرس یک پوشه با نام input شامل تعداد نسبتاً زیادی عکس با فرمت‌های مختلف یک پوشه جدید با نام processed ایجاد نماید که شامل خواص زیر باشد:

- 1 - عکس با تصویر تکراری در پوشه processed وجود ندارد.
- 2 - نام عکس های پوشه processed به شکل زیر اصلاح شده باشد:
YYYYMMDD-iiii-filename.ext
که در آن YYYYMMDD تاریخ شمسی creationDate مربوط به عکس می باشد.
-iiii یک شمارنده چهار رقمی است که از 0001 شروع می شود و برای عکس های گرفته شده در یک تاریخ unique میباشد.
- Filename نامه فعلی فایل در پوشه input میباشد که در آن کاراکترهای عدد و حروف ویژه مانند - _ ؛ () ار نام فایل حذف شده باشند.
Ext پسوند فعلی فاید در پوشه input میباشد.
- برای مثال یک نام فایل معتبر با فرضی که نام فایل اصلی در پوشه input مقدار qeshm1231-23.jpg باشد مقدار زیر میتواند باشد:
13990103-0005-qeshm.jpg
- 3 - در فایل‌های پوشه input نباید تغییری ایجاد گردد و یا پاک شوند.
- 4 - چنانچه پوشه **input یافت نشد** فانکشن process باید **خطای مناسب** ایجاد نماید و اعلام نماید آدرس پوشه ورودی صحیح نمیشود.
- 5 - **خروجی فانکشن process باید یک string حاوی آدرس پوشه processed باشد.**
- 6 - فانکشن process باید مطلب زیر را در خروجی چاپ نماید:
Number of detected input files: N, number of processed output files: n
که در آن N مقدار تعداد کلیه فایل‌های موجود در پوشه های زیر پوشه input و n تعداد فایل‌های یکتای شناسایی شده میباشد.

لازم بذکر است فایلها در پوشه input امکان دارد درون زیر پوشه های مختلف باشند و لازم است در خواندن فایلها بصورت ریکرسیو کلیه فایل‌های درون زیر فلدرهای تو در تو شناسایی شده و بصورت بالا و فلت کنار هم زیر فلدر processed ذخیره گردند.

برای مقایسه محتوای فایلها حتما از تابع hash استفاده گردد تا برنامه نوشته شده فضای حافظه قابل مدیریتی را استفاده نماید. در صورتی که تابع hash محتوای باینری دو عکس بایکدیگر یکی بود برای اطمینان از برابر بودن محتوای فایلها باید محتوای کامل و هش نشده باینری آنها را نیز با یکدیگر مقایسه نمود. اما چنانچه مقدار هش دو فایل تصویر با یکدیگر برابر نبود میتوان مطمئن بود که این دو فایل با هم اختلاف دارند.

نمره دهی از 20

4 نمره اضافی در صورت ارسال جواب قبل از نیمه شب 1401/01/17

2 نمره جهت رعایت دستخط مناسب و اصول قوانین کد نویسی پایتون (zen)

2 نمره جهت توضیح فارسی مناسب نحوه کد نویسی و کارهای انجام شده بصورت ۱ - ۲ پاراگراف در یک فایل تکست جداگانه

16 نمره جهت عملکرد ۱۰۰ درصد منطبق با توضیحات درخواستی

برای ارسال جواب تمرین باید یک ایمیل شامل یک فایل زیپ شامل موارد زیر به آدرس babakbh@gmail.com تا تاریخ گفته شده ارسال گردد. عنوان ایمیل: CDA-MIDTERM-StudentCode باشد که در آن StudentCode شماره دانشجویی شما خواهد بود.

در متن ایمیل نام نام خانوادگی و شماره دانشجویی ارسال گردد.

محتوای فایل زیپ پیوست:

1 - فایل با پسوند .py و یا ipynb شامل متن فانکشن process

2 - فایل pdf شامل توضیحات

بدیهی است تنها فایلهای قابل اجرای فرستاده شده تا تاریخ یاد شده در نظر گرفته خواهند شد.