フィックスポイント社 プログラミング試験問題

早稲田大学 服部晃平 k_hattori@aoni.waseda.jp

← → + Q = ≥ □

使用言語:Python3

開発環境:Anacondaディストリビューション下で開発

以下のライブラリを使用していますが、 全てAnaconda内に標準で含まれています。

> import numpy as np import pandas as pd import glob import matplotlib.pyplot as plt import datetime

使用方法:main.pyファイルを実行します。コードはこの1ファイルのみです。

コードを実行すると、main.pyファイルと同じディレクトリに、リモートホスト順(remotehost.csv)および時間帯ごと(timezone.csv)のアクセス件数を記載したcsvファイルを作成します。

同時に、アクセス件数を棒グラフして示した図がポップアップ表示される

ようになっています(右図参照)。

プログラムの構造について:

大容量のファイルを処理できるよう、指定した 行数ごとにファイルを読み込んで処理するように なっています。

一度に読み込む行数は、lump_sizeで設定できます。 デフォルトでは5を設定してるので、5行ずつ読み込 んでファイルを処理します。



read_pathで設定します(デフォルトでは ./var/log/httpd/*_log)。

検索範囲の指定について:

プログラム冒頭の変数で設定します。例えば、以下の場合、 $2005/4/10 \sim 2017/4/30$ の範囲を解析します。

year_start = 2005 month_start = 4 day_start = 10 year_end = 2017 month_end = 4 day_end = 30