

「Pythonとデータ分析で初学者から中級者になるためのセミナー」2023年12月20日



Pythonエンジニアになるためのテクニカルハイライトと学習方法

—Pythonを業務で使ってみよう—

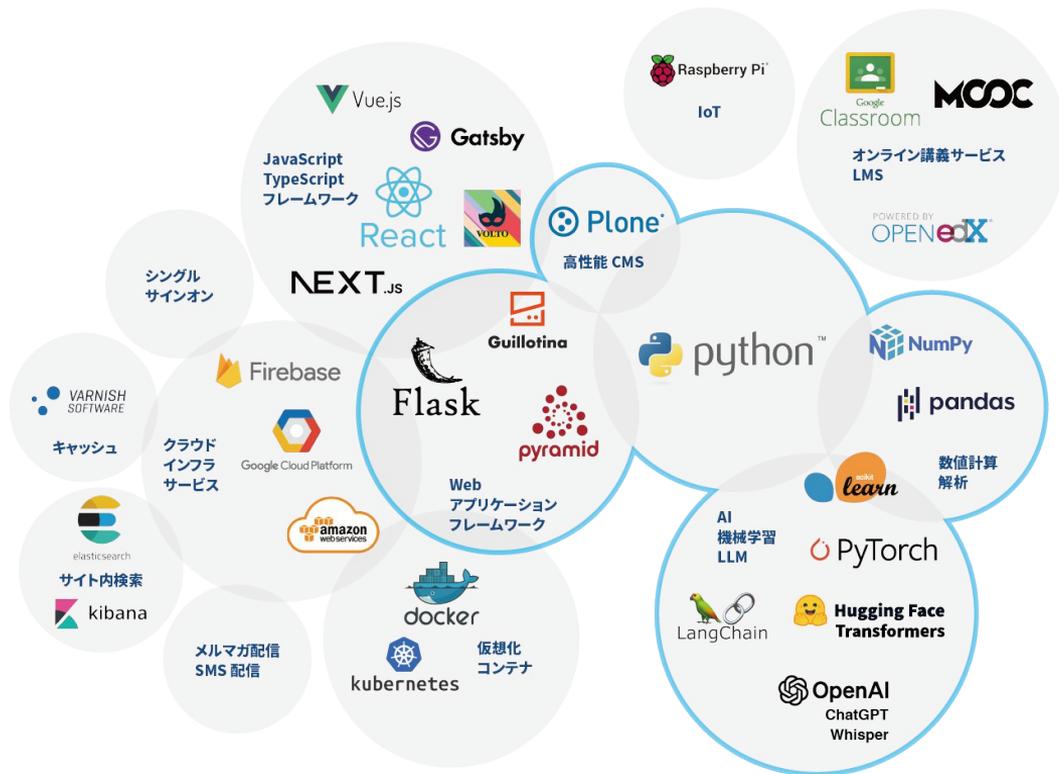
一般社団法人Pythonエンジニア育成推進協会
顧問理事 寺田 学

所属組織

- 一般社団法人Pythonエンジニア育成推進協会顧問理事
- 一般社団法人PyCon JP Association理事
- 株式会社 CMSコミュニケーションズ 代表取締役
- Python Software Foundation Fellow
- Plone Foundation Ambassador

Python技術支援のコンサルサービス

(株)CMSコミュニケーションズのサービス



講演の流れ

- Pythonバージョンの現状
- バージョンアップで取り組んでいること
- 用途別定番ライブラリ
- ライブラリの具体的な利用シーン
- 初学者からアプリ実装ができるようになるには

Pythonバージョンの現状

Pythonバージョン(2023年12月)

Python3.12

- Python 3.10以降を採用することが多い
- 最新のPython3.11はようやく使い始めている

バージョンアップで取り組んでいること

バージョンアップで取り組んでいる(1)

- スピードアップ
- Python 3.11から取り組んでいる
- 条件によっては1.5倍早くなっている
- 今後も内部実装を見直し、スピードアップに期待
- GIL関係のリファクタリング

用途別定番ライブラリ

用途別定番ライブラリー - 3

Webシステム

- Webフレームワーク Django / Flask / FastAPI
- UIフレームワーク(データ分析可視化) Streamlit / Gradio
- O/Rマッパー SQLAlchemy

ライブラリの具体的な利用シーン

具体的な利用シーン - 1

ログの分析をPythonで行う

- ログファイルをPythonのopen関数で読み込む
- 1行の中身を分割して、リストのリストを作る
- pandasで読み込み
- 必要な情報だけを抽出
- 可視化やExcelファイルに書き出し
- 機械学習などを用いて、異常検知

具体的な利用シーン - 1

ログの分析で使うライブラリ

- 日付操作 標準ライブラリ `datetime`
- データ分析 `pandas`
- データ出力 `openpyxl`
- 可視化 `Matplotlib`
- 機械学習 `scikit-learn`

具体的な利用シーン - 2

定期的にCSVファイルを取得してデータ登録

- CSVファイルを取得
- データを確認してDBにデータ登録
- 完了時にメールを送信する
- エラー時はアラートを送信する

具体的な利用シーン - 2

CSVファイルからデータ登録で使うライブラリ

- CSVファイル 標準ライブラリ `csv`
- DBへの登録 `SQLAlchemy`
- メール送信 標準ライブラリ `smtplib`

具体的な利用シーン - 3

データ分析のPoCを試してみたい

- JupyterLabで実験的に試行錯誤
- データ加工
- データの評価
- 各種Webサイトから情報取得
- 結果をWebページにする

初学者からアプリ実装ができるようになるには

機能分割

まずは、機能を分割していこう

- アプリが実現したいことを機能分割
- 例えば
 - データを読み込む
 - データを加工する
 - 画面に表示する
- なるべく細かく分割する

機能を繋げる

機能を繋げて、アプリとしての動きを実現

- 機能を繋げて、一連の動きにする
- 何かしらのキッカケで動作するようにする
- キッカケの例
 - クリックする
 - スクリプトの手動起動
 - Webリクエスト

アプリ実装のまとめ

一つずつ確実に!

- 機能分割
- 機能ごとに実装
- 機能を繋げる
- 動作環境の整備

師匠 or 仲間を見つけよう

ひとりで勉強するには限界がある

- 師匠を見つけよう
- 仲間を見つけよう
- 一緒に勉強し切磋琢磨できるとよい
- 自分のやっていることに自信が持てる
- カンファレンスや勉強会に積極的に参加

PyCon APAC 2023

約700人規模の現地イベント

- <https://2023.pycon.jp/>
- 日程: 10/26(金)~28(日) 開催済み
- 2024も秋にPyCon JP開催予定
- 当協会はGoldスポンサー



Q&A