環境構築マニュアル

１Dockerのインストール

こちらのウェブ記事を参照：<https://lab.sonicmoov.com/development/install-docker-windows/>

・「BIOSでVirtualization Technologyを有効化」いらないはず

2 Dockerコマンド

1. Dockerhubが起動しているのを確認

➁　ベースイメージを検索：docker search [OPTIONS]

e.g. docker search python:3.8

③必要なイメージをダウンロード：docker image pull [NAME]

e.g. docker image pull python:3.8

④イメージの一覧表示をする：docker image ls

⑤コンテナに入る: docker run -it [NAME] /bin/bash

⑥エディタインストール：apt-get update

apt-get install -y vim

こちらのウェブ記事を参照：<https://qiita.com/gold-kou/items/44860fbda1a34a001fc1>

3 VScode

1. VScodeインストール

こちらのウェブ記事を参照：<https://qiita.com/psychoroid/items/7d85ae6bade4a67aedb1>

1. VScodeの「extensions」で「Remote-Containers」を検索、インストールする。
2. 左下角の「><」をクリックして表示されるポップアップメニューから、「Remote-Containers: Open Folder in Container…」をクリックしてプロジェクトのフォルダを開く
3. 「Attach to Running Container…」を選択する

**これでVScodeがdocker環境と接続成功。**

**VScode→docker接続→「NewTerminal」でファイル実行やモジュールインストール**

**コマンドプロンプト→dockerに入る→ファイル実行やモジュールインストール**

４Scrapyインストール

元データからScrapyを用いてPDFをダウンロード

pip install scrapy

5 Apache Tika

PDFをテキスト化

pip install tika

Dockerfile調査

<https://qiita.com/Riliumph/items/3b09e0804d7a04dff85b>