

YIC 情報ビジネス専門学校

チーム クローラ.py

MBSD Cybersecurity Challenges

泳げ!簡単かいせきクン説明書

目次

1.はじめに.....	2
2.ツールの優位性.....	3
3.機能説明	4
4.インストール手順.....	6
5.操作説明書.....	11

1 .はじめに

泳げ!簡単かいせきクンはラズパイを使ったクローラーです。

ラズパイに SD カードを差し込むだけでクロール環境が作れる「簡単さ」がうりです。

なお、本説明書は泳げ!簡単かいせきクンのインストールおよび取り扱いについて説明するもので ある。

推奨環境

Raspberry Pi 4 model B

windows11

2. ツールの優位性

▶ インストールが単純！

用意するものはラズパイだけです！送付した SD カードを指して、起動するだけでクローलする準備できます！

※今回は、メールでの提出なのでラズパイのイメージファイルを SD カードにインストールする形になります。

▶ Web 診断が楽に！

クロール結果に、脆弱性の可能性があるページが出力されます！

クロール後の Web 診断では、そのページを注力して診断することができ、効率的に進めることができます！

3.機能説明

3.1 基本機能

・クローリング機能

巡回を開始する URL を入力すると、そのページからたどれる範囲を巡回して、表にリストアップして出力します。

リストアップの内容

クローラーの基本的な出力	脆弱性診断用の出力
URL	cookie の有無
タイトル	入力フォームの有無
パラメータ	ダウンロード、アップロードの有無
キーワード※	メールフォームの有無
タイムスタンプ	ログイン処理の有無
	html の byte 数

※キーワードは、テストサイトのコンテンツ内に、MBSD{xxxxxx} (カッコ内は任意の英数字) があった場合にカッコ内を抽出して出力します。

・クローリング履歴機能

クローリング結果を保存が履歴として残ります。詳細ボタンを押すと、クローリング時と同じ結果を見ることができます。

・ソート機能

出力された表の各カラムをソートすることができます。

・ログイン画面に対応

泳げ!簡単かいせきクンの入力フォームにユーザ名とパスワードを入力することでログイン後の画面もクローリングできます。

・ローカル環境での実行が可能

開発環境で使用することを想定し、LAN 内で (インターネットに接続されていなくても) 使用出来ます。

3.2 詳細機能

- ・ツール実行者への補助機能

- ・Q & A の実装
- ・操作手順書の実装

- ・バリデーション機能

使いやすさを考え、URL とドメインの入力フォームにバリデーション機能を入れています。
間違った値を入力して、クロールボタンを押しても、エラーを返します。

- ・タイムアウトに対応

ロードに 10 秒以上かかる Web ページは飛ばしてクロールします。

4.インストール手順

※本手順書はエンジニアの方を対象にしています。
基本的なネットワークの知識が必要になる箇所があります。

4.1 用意するもの

表 4.1 必要なもののリスト

	必要なもの	説明
1	ラズベリーパイ 4 本体	Raspberry Pi 4 Model B
2	Micro SD カード	送付した SD カード
3	USB(Type-C) 電源ケーブル	
4	Micro HDMI ケーブル	
5	モニター	ラズベリーパイ用のモニター
6	ラズベリーパイ用ケース	なくても大丈夫です。
7	USB キーボード	ラズベリーパイ用のキーボード
8	USB マウス	ラズベリーパイ用のマウス
9	Lan ケーブル	
10	パソコン	ラズベリーパイの Web サーバにアクセスする。

4.2 ラズパイイメージの用意

ラズパイ OS をインストールした SD カードを送付する予定でしたが、
メールでの提出のためイメージファイルをダウンロードして、SD カードにインストールして頂きます。

4.2.1 Raspberry Pi Imager のインストール

Raspberry Pi Imager は公式が用意しているラズパイ OS をイメージに書き込めるツールです。

以下の URL から windows 版をインストールしてください。

<https://www.raspberrypi.com/software/>

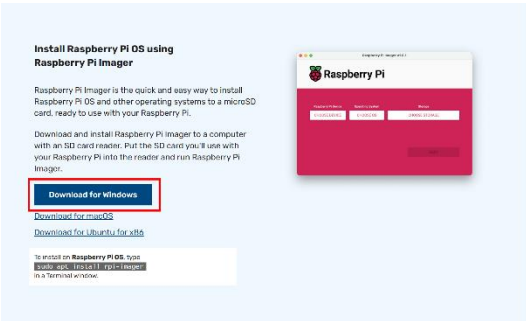


図 4.1 Raspberry Pi Imager インストールページ

「Download for Windows」をクリックしてください。

「Install」のクリックでインストールが始まります。

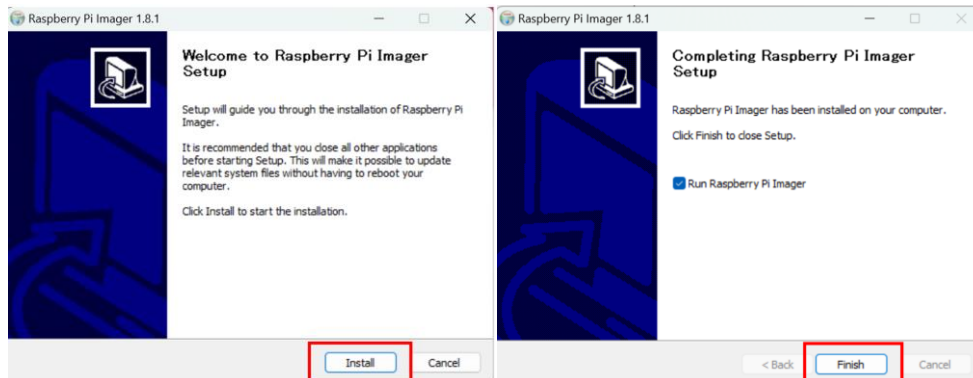


図 4.2 Raspberry Pi Imager のインストール画面

「Finish」をクリックして、以下の画面が出たら成功です。



図 4.2 Raspberry Pi Imager

4.2.2 イメージファイルをダウンロード

イメージファイルを以下の URL からダウンロードします

※ クリックするとダウンロードが始まります。

また、ファイルの容量は約 3Gbyte です。

<https://www.shonanwind.com/mbsd/kaiseki/backup.img.gz>

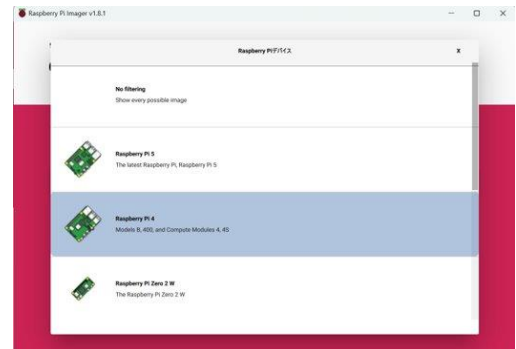
ハッシュ値を使ってファイルの同一性を確認できます。

詳しくは Hash ディレクトリの中を参照

4.2.3SD カードにイメージファイルを書き込む

デバイスを選択

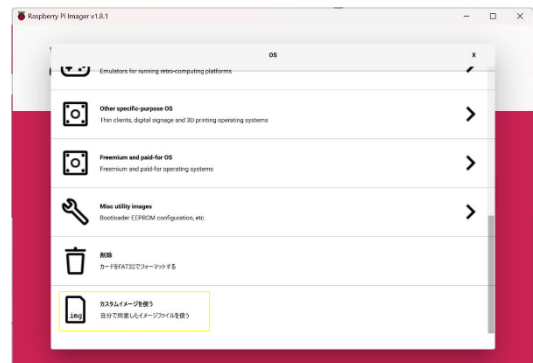
「デバイスを選択」をクリックして、Raspberry Pi 4 を選択



OS を選択

「OS を選択」をクリックして、「カスタムイメージを使う」に進む。

3.2.2 でダウンロードしたイメージファイル「backup.img.gz」を選択

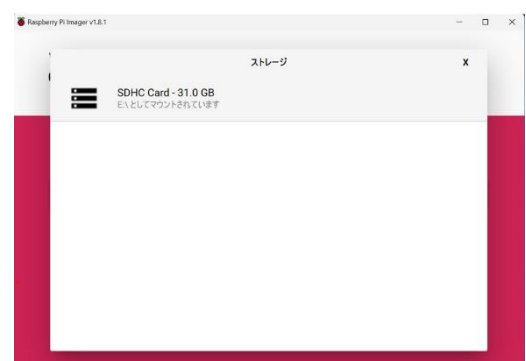


ストレージを選択

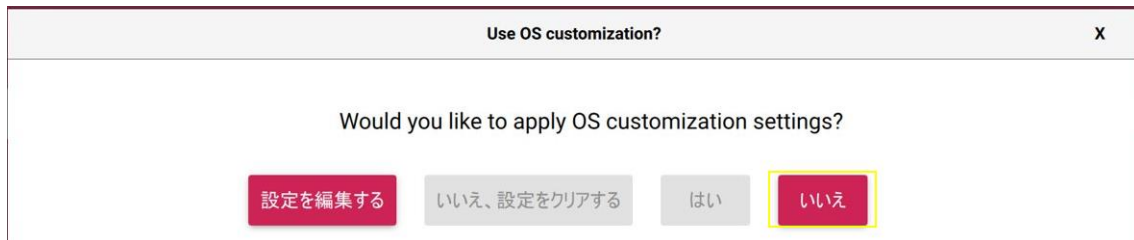
パソコンに 32Gbyte 以上の SD カードを差し込む

「ストレージを選択」をクリックして、

差し込んだ SD カードを選択



「次へ」、「いいえ」、「はい」とクリックする。



SD カードの書き込みは 10 分程度かかります。

4.2 ラズパイの起動

- 1 ラズパイに MicroSD カード、HDMI ケーブル、キーボード、マウス、Lan ケーブルを接続する。
- 2 電源ケーブルを接続するとモニターにラズベリーパイの起動画面が表示される
- 3 ログイン

ユーザ名 : yoshimatsu

パスワード : 08yoshi09

4.3 固定 IP アドレスの設定

※初期設定は DHCP で IP アドレスを振る設定になっています。

\$ ip a コマンドを使い、ラズパイの IP アドレスを調べて Web アクセスすることも可能です。インストールを簡単にしたい場合、固定 IP アドレスの設定は必要ありません。

4.4 設定項目を調べる

4.4.1 以下を調べて下さい。分からない場合、ネットワーク管理者へ相談してください。

- ・使用しても良い ip アドレス
- ・サブネットマスク
- ・デフォルトゲートウェイの IP アドレス
- ・DNS サーバの IP アドレス

4.4.2/etc/network/interfaces の編集

```
$ sudo vi /etc/network/interfaces
```

```
source /etc/network/interfaces.d/*
```

```
auto lo
```

```
iface lo inet loopback
```

```
auto eth0
```

```
iface eth0 inet static
```

```
address 使用してもいい IP アドレス
```

```
netmask サブネットマスク
```

```
gateway デフォルトゲートウェイ
```

```
dns-nameserver DNS サーバ
```

4.4.3 設定を適応

```
$ sudo systemctl restart networking
```

4.4.4 設定が変更されたか確認

```
$ ip a コマンドを使い、ip アドレスが変更されていることを確認
```

4.4.5 ブラウザアクセスが出来るか確認

ラズベリーパイと同一ネットワークに接続しているパソコンでブラウザを開き、
ラズベリーパイの IP アドレス:8000 で接続確認

5.操作説明書

5.1 泳げ!簡単かいせきクンのトップ画面を表示する

http://4 で確認、または設定した IP アドレス : 8000/

次のような、泳げ!簡単かいせきクンのトップ画面が表示されると成功です。

※図 5-1 はクロール履歴がありますが、初期段階ではクロール履歴がありません。

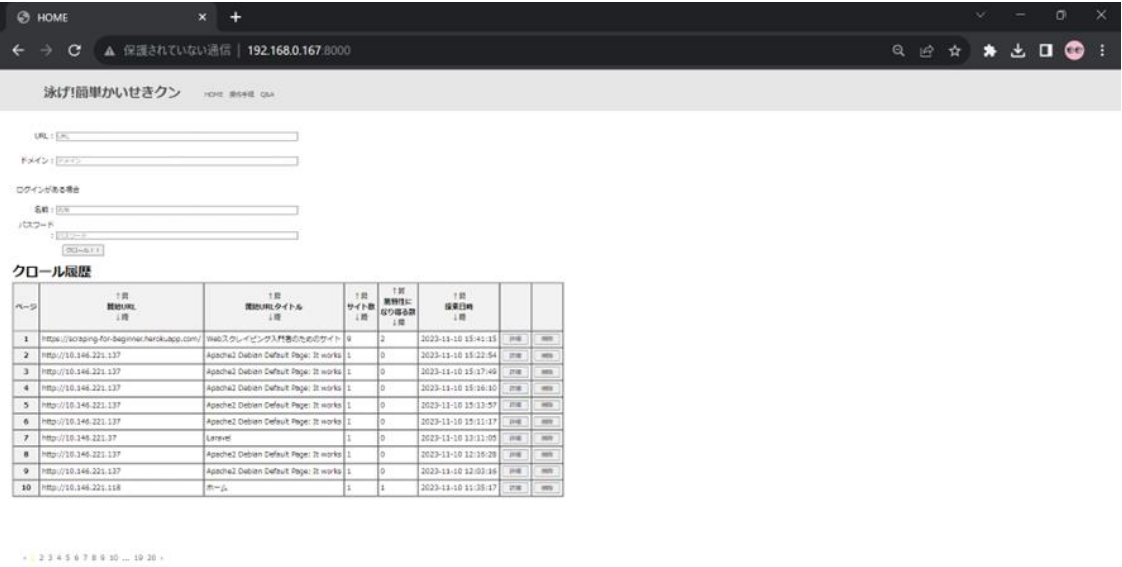


図 5.1 泳げ!簡単かいせきクン トップ画面

5.2 クロールの実行方法

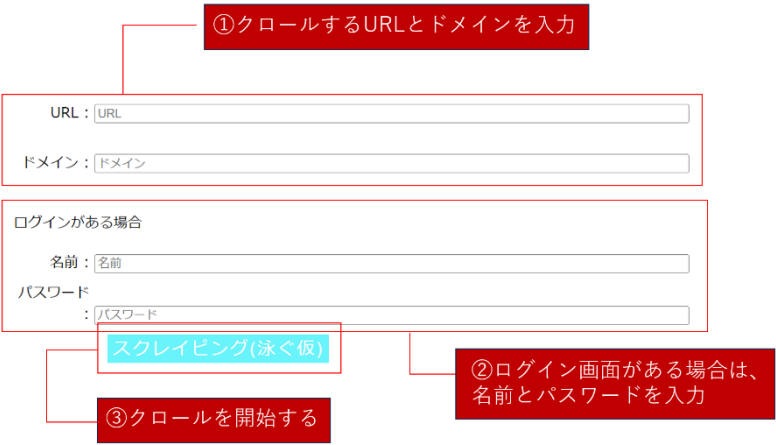


図 5.2.1 泳げ!簡単かいせきクン 入力フォーム

クロールが終了するとクロール結果画面に遷移します。

泳げ!簡単かいせきクン HOME 操作手順 Q&A

クロール結果

#	URL	タイトル	パラメータ	キーワード	Cookie	入力 フォーム	アップロード ダウンロード	ページ容量 (byte)
1	https://scraping-for-beginner.herokuapp.com	Webスクレイピング入門者のためのサイト			x	x	x	4076
2	https://scraping-for-beginner.herokuapp.com/	Webスクレイピング入門者のためのサイト			x	x	x	4076
3	https://scraping-for-beginner.herokuapp.com/image	Webスクレイピング入門者のためのサイト			x	x	x	5079
4	https://scraping-for-beginner.herokuapp.com/login_page	Webスクレイピング入門者のためのサイト			○	○	x	3264
5	https://scraping-for-beginner.herokuapp.com/mypage	Webスクレイピング入門者のためのサイト			x	x	x	2798
6	https://scraping-for-beginner.herokuapp.com/ranking/	Webスクレイピング入門者のためのサイト			x	x	x	17175
7	https://scraping-for-beginner.herokuapp.com/ranking/?page=1	Webスクレイピング入門者のためのサイト	page:1		x	x	x	17224
8	https://scraping-for-beginner.herokuapp.com/ranking/?page=2	Webスクレイピング入門者のためのサイト	page:2		x	x	x	17171
9	https://scraping-for-beginner.herokuapp.com/ranking/?page=3	Webスクレイピング入門者のためのサイト	page:3		x	x	x	17108
10	https://scraping-for-beginner.herokuapp.com/udemy	Webスクレイピング入門者のためのサイト			x	x	x	3166

< 一覧画面へ戻る csv出力

data.csv
1,303 B • 完了

図 5.2.2 クロール結果画面の説明

※https://scraping-for-beginner.herokuapp.com にクロールしています。

5.3 クロール履歴の確認

泳げ!簡単かいせきクン HOME 操作手順 Q&A

URL:

ドメイン:

ログインがある場合
名前:

パスワード:

クリックすると選んで昇順、降順でソートできます

クロール履歴

ページ	開始時刻	開始時刻/タイトル	サイト数	最終的に 残った数	履歴日時	削除
1	2023-11-12 22:45:03	Webスクレイピング入門者のためのサイト	10	1	2023-11-12 22:45:03	削除
2	2023-11-12 22:43:21	Webスクレイピング入門者のためのサイト	10	2	2023-11-12 22:43:21	削除
3	2023-11-12 22:42:38	Webスクレイピング入門者のためのサイト	1	0	2023-11-12 22:42:38	削除
4	2023-11-10 15:41:15	Webスクレイピング入門者のためのサイト	9	2	2023-11-10 15:41:15	削除
5	2023-11-10 15:22:54	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:22:54	削除
6	2023-11-10 15:17:49	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:17:49	削除
7	2023-11-10 15:16:10	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:16:10	削除
8	2023-11-10 15:14:39	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:14:39	削除
9	2023-11-10 15:13:57	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:13:57	削除
10	2023-11-10 15:11:17	Apache2 Debian Default Page: It works	1	0	2023-11-10 15:11:17	削除

1/2 168.5.167.8000 の内蔵
ホストによるものです

OK キャンセル

クリックすると過去の
クロール結果の詳細を確認できます。

ページネーション

図 5.3 クロール履歴の説明

5.4 ヘッダーの説明



図 5.4 ヘッダーの説明