

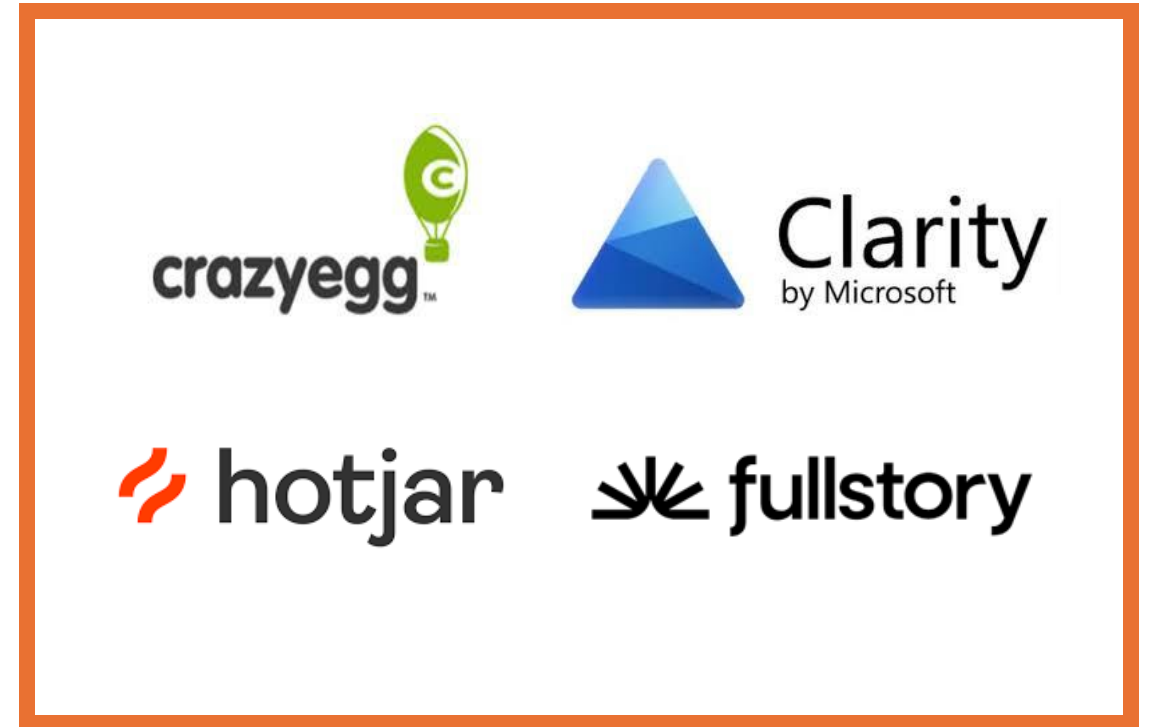
ウェブサイトからマウス履歴を取得する セッションリプレイサービスの検出ツールの開発

田口凱之, 菊池浩明

明治大学

セッションリプレイサービスとは

- セッションリプレイサービス
 - ユーザーのWebサイト上での操作
 - マウスの移動
 - クリック
 - スクロール
 - ページ閲覧情報
 - デバイス情報などを記録するサービス
 - ユーザ体験の向上やサイト改善

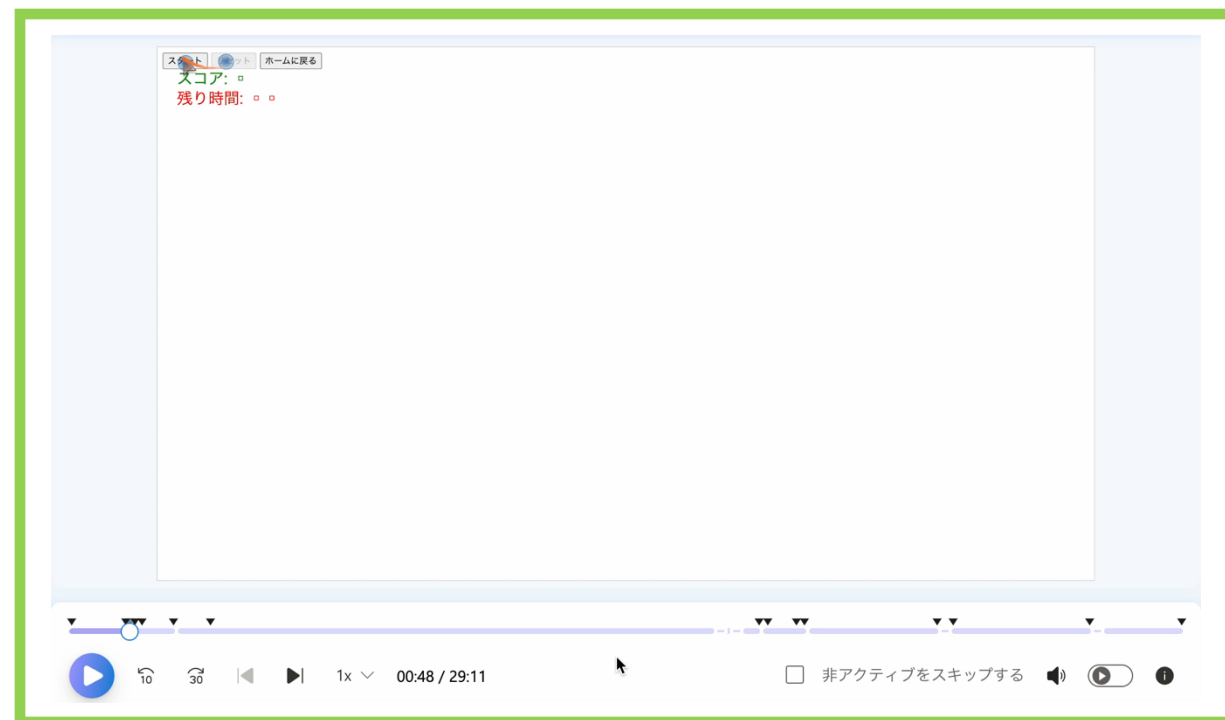


セッションリプレイデモンストレーションサイトの開発

デモサイト



Clarity のモニタ



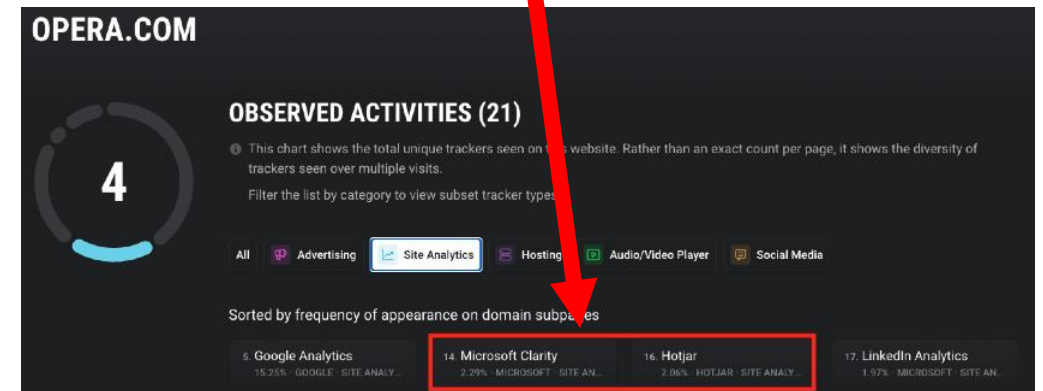
目的

- **ウェブサイトに導入されているセッションリプレイサービスを自動的に検出するツールを開発する**
- 主要なウェブサイトを対象にツールの検出精度を明らかにする

セッションリプレイサービス導入状況調査

- 調査目的
 - セッションリプレイサービスの導入率
 - どのサービスが主流か
- 調査対象
 - **Tranco**の2024年度データセット
 - 世界の上位70サイト
 - 日本の上位70サイト
 - **10種類**のセッションリプレイサービス
- 調査方法
 - **Ghostery** ;
トラッキング技術を検出・排除する
ブラウザ拡張ツール

検出したセッションリプレイサービス



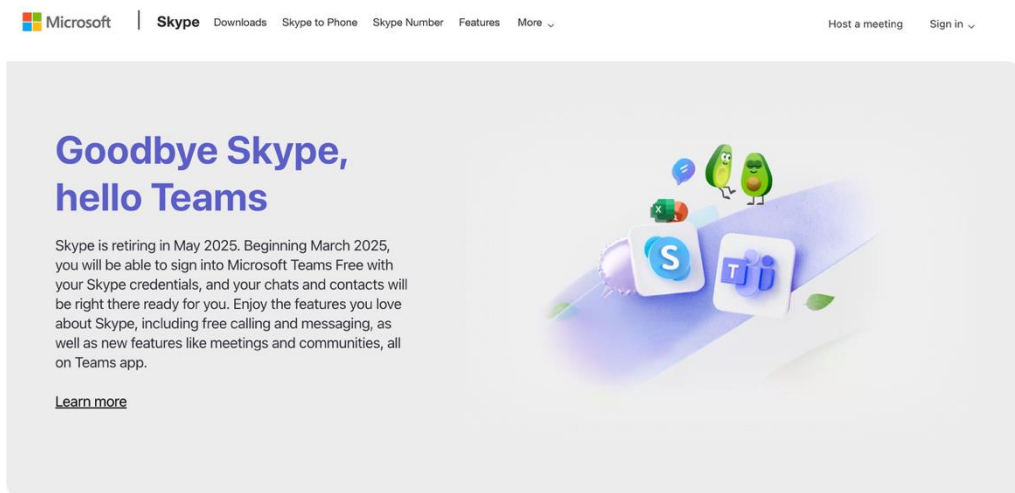
Ghosteryの動作画面

調査結果

サービス名	世界（70サイト）		日本（70サイト）	
	数	%	数	%
Microsoft Clarity	12	17.1	12	17.1
Hotjar	2	2.9	6	8.6
Mouseflow	0	0.0	0	0.0
Yandex Metrica	2	2.9	0	0.0
Contentsquare	0	0.0	0	0.0
Crazyegg	1	1.4	1	1.4
Dynatrace	0	0.0	1	1.4
FullStory	1	1.4	3	4.3
LogRocket	0	0.0	0	0.0
Lucky Orange	0	0.0	0	0.0
合計	18	25.7	23	32.9

セッションリプレイサービス検出原理

Webサイト



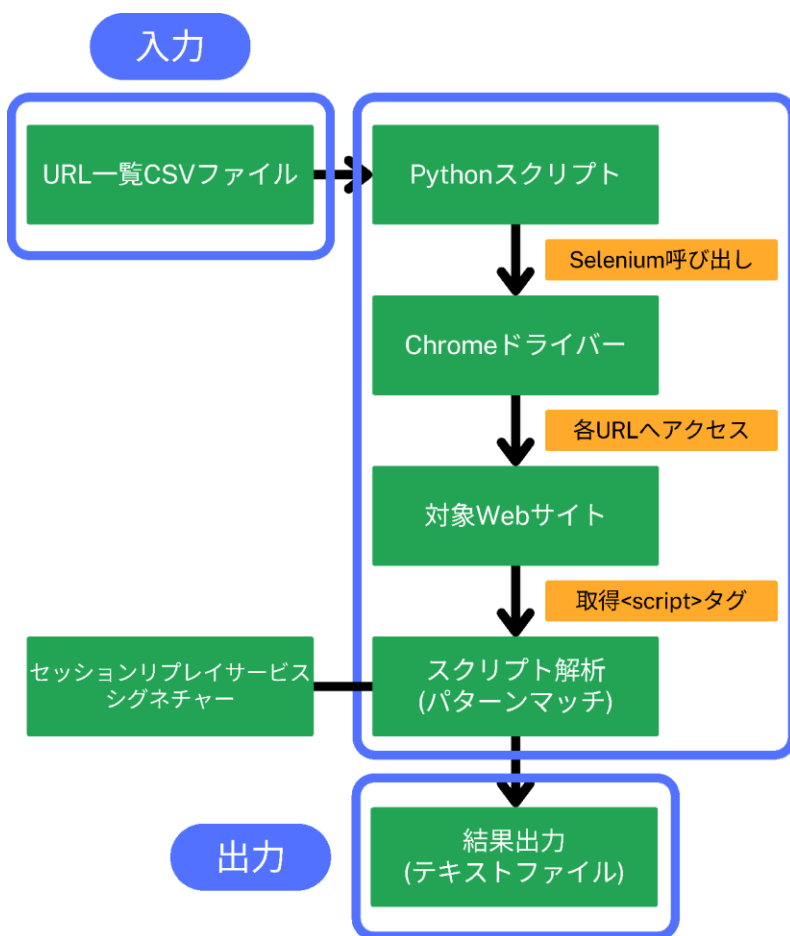
ソースコード(一部)

```
<script type="text/javascript">
  (function (c, l, a, r, i, t, y)
  { c[a] = c[a] || function ()
  { (c[a].q = c[a].q || []).push(arguments) };
  t = l.createElement(r); t.async = 1;
  t.src = https://www.clarity.ms/tag/+i;
  y=l.getElementsByTagName(r)[0];
  y.parentNode.insertBefore(t,y);
  (window,document,"clarity","script",c1i4yeb6q8");
</script>
```

セッションリプレイサービス固有のコードパターン
Microsoft Clarityの例

検出ツールの開発

検出アルゴリズム



実行例

```
-----  
Checking URL: https://www.tp-link.com/jp/  
-----
```

```
Detected session replay services:  
- Crazy Egg  
-----
```

```
Checking URL: https://www.xbox.com/ja-JP/live  
-----
```

```
Detected session replay services:  
- Microsoft Clarity  
-----
```

```
Checking URL: https://www.xserver.ne.jp/  
-----
```

```
No session replay services detected.  
-----
```

```
Checking URL: https://www.yahoo.co.jp/  
-----
```

```
No session replay services detected.
```

検出結果

世界の上位70サイト

真\検出	検出あり	検出なし
セッションリプレイあり	10	8
セッションリプレイなし	3	49

日本の上位70サイト

真\検出	検出あり	検出なし
セッションリプレイあり	7	16
セッションリプレイなし	2	45

検出精度

	Precision	Recall	F1
世界	0.769	0.556	0.645
日本	0.778	0.304	0.438
平均	0.774	0.43	0.542

FP

FN

未検出・誤検出の原因

• 未検出 (FN)

- <https://www.bing.com/>や<https://unity.com/j>など
- トップページソースコードに定義済みのパターンなし
- トップページ以外にセッションリプレイサービスが導入されている可能性
- 異なるコードパターンやURLを用いている可能性

• 誤検出 (FP)

- <https://wordpress.com/ja/>
- 定義済みのパターンと一致する文字列 (clarity.ms/tag/)を確認
- 以前使われており現在は動作していない可能性

view-source:https://wordpress.com/ja/

```
window.clarity =  
  window.clarity ||  
  function () {  
    ( window.clarity.q = window.clarity.q || [] ).push( arguments );  
  };  
  
const clarityScript = kit.attachScriptElement( 'https://www.clarity.ms/tag/j0cc1i1dba' );  
document.body.appendChild( clarityScript );
```

wordpress.comのソースコードの一部

結論

- 国内外の主要なウェブサイトにおけるセッションリプレイサービスの導入状況を調査し、**25%(世界)**、**32%(日本)**の利用率を明らかにした
- セッションリプレイサービス検出ツールを開発し、平均F1で**0.542**の精度を評価した
- 今後の課題としてユーザにセッションリプレイサービスの存在を分かりやすく伝える手法の開発を挙げる

参考文献

1. 梶間大地, 菊池浩明, “セッションリプレイサービスからの個人識別性と国内外サイトにおけるプライバシーポリシーでの公表状況”, 第104回CSEC研究発表会, 2024-CSEC-104, No.11, pp1-8, IPSJ, 2024.
2. Selenium, (<https://www.selenium.dev/ja/documentation/>, 2024年8月参照).
3. Ghostery, (<https://www.ghostery.com/>, 2024年8月参照).
4. Tranco, (<https://tranco-list.eu/>, 2024年8月参照).