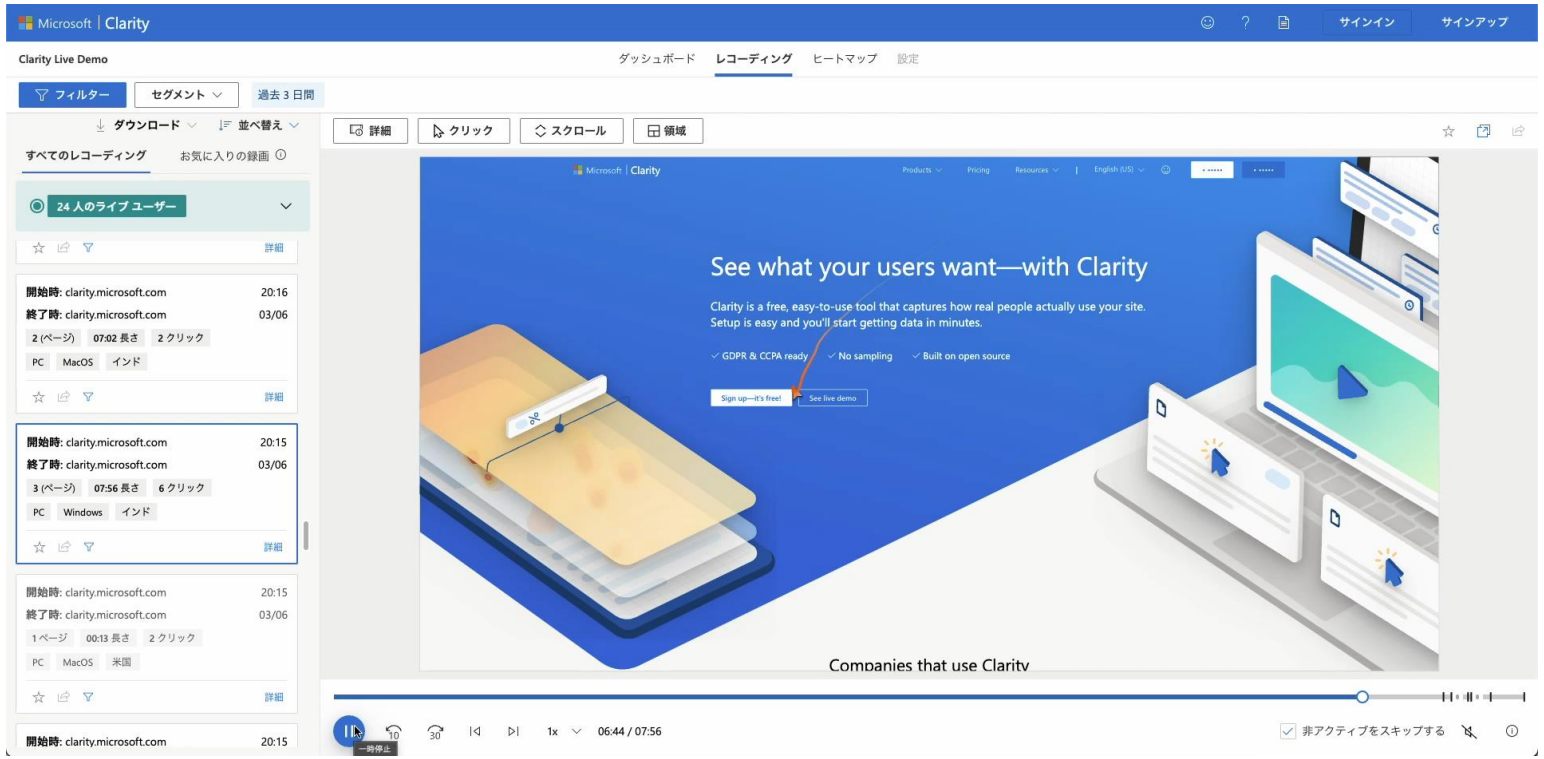
セッションリプレイサービスからの個人識別性と国内外サイトにおけるプライバシーポリシーでの公表状況調査

梶間大地　菊池浩明

明治大学

セッションリプレイサービスとは



Microsoft Clarityデモ

ウェブサイトの改善ツール

個々のユーザの操作を記録

利用目的

- UX改善
- エラー特定

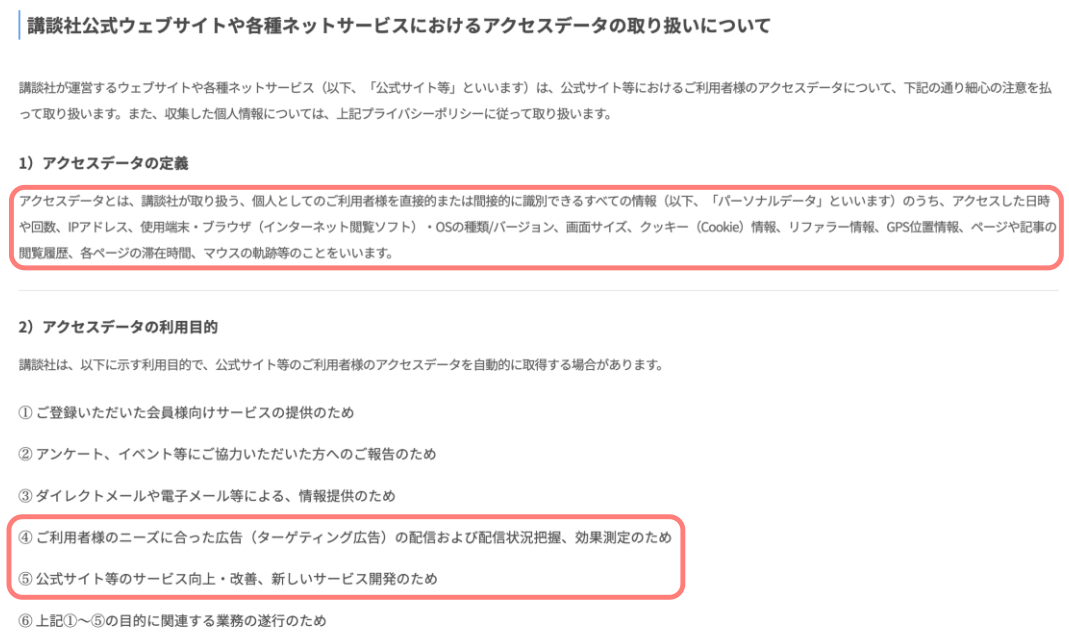
Clarity by Microsoft

hotjar

mouseflow

セッションリプレイサービスの問題点

1. 事業者が導入をポリシーで公表していない場合がある
2. 導入しているかどうかが利用者が気付けない
3. マウス・キーボードの操作履歴から個人が識別されるリスク



講談社公式ウェブサイトや各種ネットサービスにおけるアクセスデータの取り扱いについて

講談社が運営するウェブサイトや各種ネットサービス（以下、「公式サイト等」といいます）は、公式サイト等におけるご利用者様のアクセスデータについて、下記の通り細心の注意を払って取り扱います。また、収集した個人情報については、上記プライバシーポリシーに従って取り扱います。

1）アクセスデータの定義

アクセスデータとは、講談社が取り扱う、個人としてのご利用者様を直接的または間接的に識別できるすべての情報（以下、「パーソナルデータ」といいます）のうち、アクセスした日時や回数、IPアドレス、使用端末・ブラウザ（インターネット閲覧ソフト）・OSの種類/バージョン、画面サイズ、クッキー（Cookie）情報、リファラー情報、GPS位置情報、ページや記事の閲覧履歴、各ページの滞在時間、マウスの軌跡等のことをいいます。

2）アクセスデータの利用目的

講談社は、以下に示す利用目的で、公式サイト等のご利用者様のアクセスデータを自動的に取得する場合があります。

① ご登録いただいた会員様向けサービスの提供のため

② アンケート、イベント等にご協力いただいた方へのご報告のため

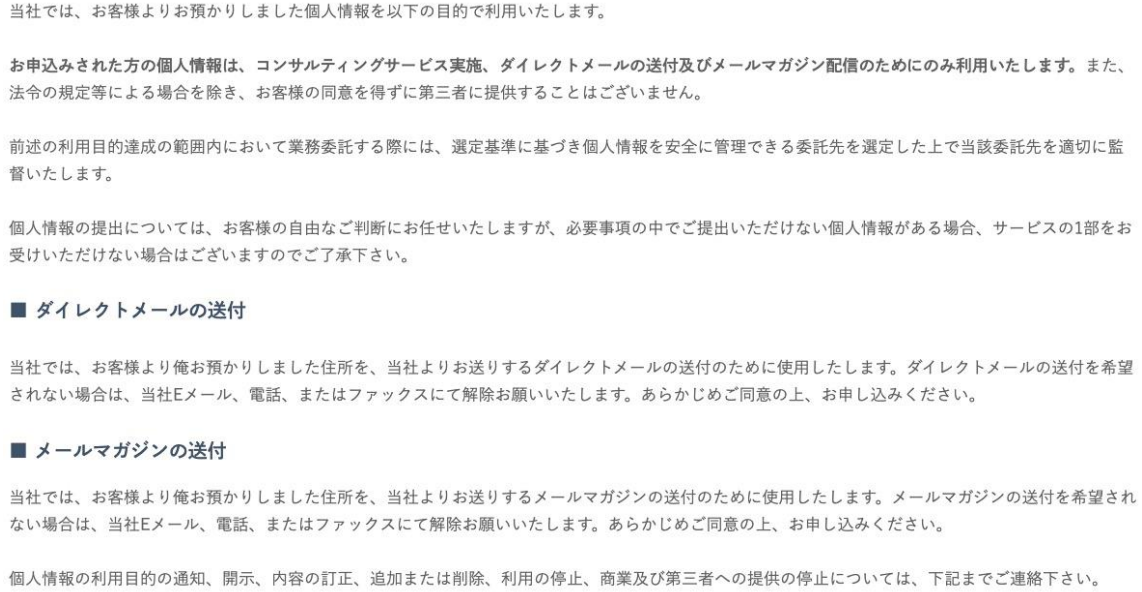
③ ダイレクトメールや電子メール等による、情報提供のため

④ ご利用者様のニーズに合った広告（ターゲティング広告）の配信および配信状況把握、効果測定のため

⑤ 公式サイト等のサービス向上・改善、新しいサービス開発のため

⑥ 上記①～⑤の目的に関連する業務の遂行のため

出版社のプライバシーポリシー[2]



当社では、お客様よりお預かりしました個人情報を以下の目的で利用いたします。

お申込みされた方の個人情報は、コンサルティングサービス実施、ダイレクトメールの送付及びメールマガジン配信のためにのみ利用いたします。また、法令の規定等による場合を除き、お客様の同意を得ずに第三者に提供することはございません。

前述の利用目的達成の範囲内において業務委託する際には、選定基準に基づき個人情報を安全に管理できる委託先を選定した上で当該委託先を適切に監督いたします。

個人情報の提出については、お客様の自由なご判断にお任せいたしますが、必要事項の中でご提出いただけない個人情報がある場合、サービスの1部をお受けいただけない場合はございますのでご了承下さい。

■ ダイレクトメールの送付

当社では、お客様より俺お預かりしました住所を、当社よりお送りするダイレクトメールの送付のために使用したします。ダイレクトメールの送付を希望されない場合は、当社Eメール、電話、またはファックスにて解除お願いいたします。あらかじめご同意の上、お申し込みください。

■ メールマガジンの送付

当社では、お客様より俺お預かりしました住所を、当社よりお送りするメールマガジンの送付のために使用したします。メールマガジンの送付を希望されない場合は、当社Eメール、電話、またはファックスにて解除お願いいたします。あらかじめご同意の上、お申し込みください。

個人情報の利用目的の通知、開示、内容の訂正、追加または削除、利用の停止、商業及び第三者への提供の停止については、下記までご連絡下さい。

通販会社のプライバシーポリシー[2]

[1]. GameWith 外部送信ポリシー, https://gamewith.jp/privacy/external_transmission_rules
[2]. プライバシーポリシー | クリームチームマーケティング, https://creamteam.jp/privacypolicy/

先行研究

- ウェブサイトのプライバシーに関する研究
  1. サードパーティのスクリプトによる情報流出について50,000サイトを調査[1]

     → 6つのセッションリプレイサービスのDOMの収集によって機密情報が流出
  2. 病院の19,483のウェブサイトのセキュリティとプライバシーを調査[2]

     → **ユーザ名，電話番号，メールアドレス**をセッションリプレイサービスに送信するサイトを確認
- 操作による個人識別研究
  - マウスの移動パターンによる個人識別[3] → 70%以上の精度で識別
  - コンピュータのスリープ復帰時の操作に着目した研究[4] → 4人ではEER2.5%，12人では10.0%で識別
  - キーボードとマウス操作ログから識別する研究[5] → ログの操作者が同一であるかを90%以上で識別
  - スマートフォンで文書閲覧時の操作から識別[6] → 複数の特徴量を用いることで識別が可能

[1]. Gunes Acar et al. “No boundaries: data exfiltration by third parties embedded on web pages”，Proceeding of the 20th Privacy Enhancing Technologies Symposium，pp.220-238，2020.
[2]. Xiufen Yu et al. “Got Sick and Tracked: Privacy Analysis of Hospital Websites”，IEEE European Symposium on Security and Privacy Workshops，pp.278-286，2022.
[3]．泉正夫，長尾若，宮本貴朗，福永邦夫，“マウス操作の特徴を用いた個人識別システム”，電子情報通信学会論文誌B，pp.305-308，2004.
[4]．須田 恭平，石田 繁巳，稲村 浩，中村 嘉隆，“日常的な家電操作による人物識別のため のマウス操作による検討．情報処理学会研究報告，モバイルコンピューティングと新社会システム研究会”，pp.1–6，2021.
[5]．木村悠生，猪俣敦夫，上原哲太郎，“深層学習を用いたキーボード入力とマウス操作情報による個人識別”，コンピュータセキュリティシンポジウム（CSS2022），pp.493-499，2022.
[6]．渡邊裕司，市川俊太，“スマートフォンにおけるタッチ操作の特徴を用いた継続的な個人識別システムの検討”，コンピュータセキュリティシンポジウム（CSS2012），pp.797-804，2012.

リサーチクエスチョン

RQ1. セッションリプレイサービスが取得する情報や利用目的に対してユーザは許容するのか？

RQ2. セッションリプレイサービスによってユーザの履歴が取得されていることに気づけるか？

RQ3. セッションリプレイサービスによる個人識別性はあるか？

RQ4. セッションリプレイサービスの導入状況

RQ5. 国内外での導入サイトの傾向はあるか？

RQ6. 導入サイトがユーザに対して適切に公表を行っているか？

ユーザ感情の調査方法

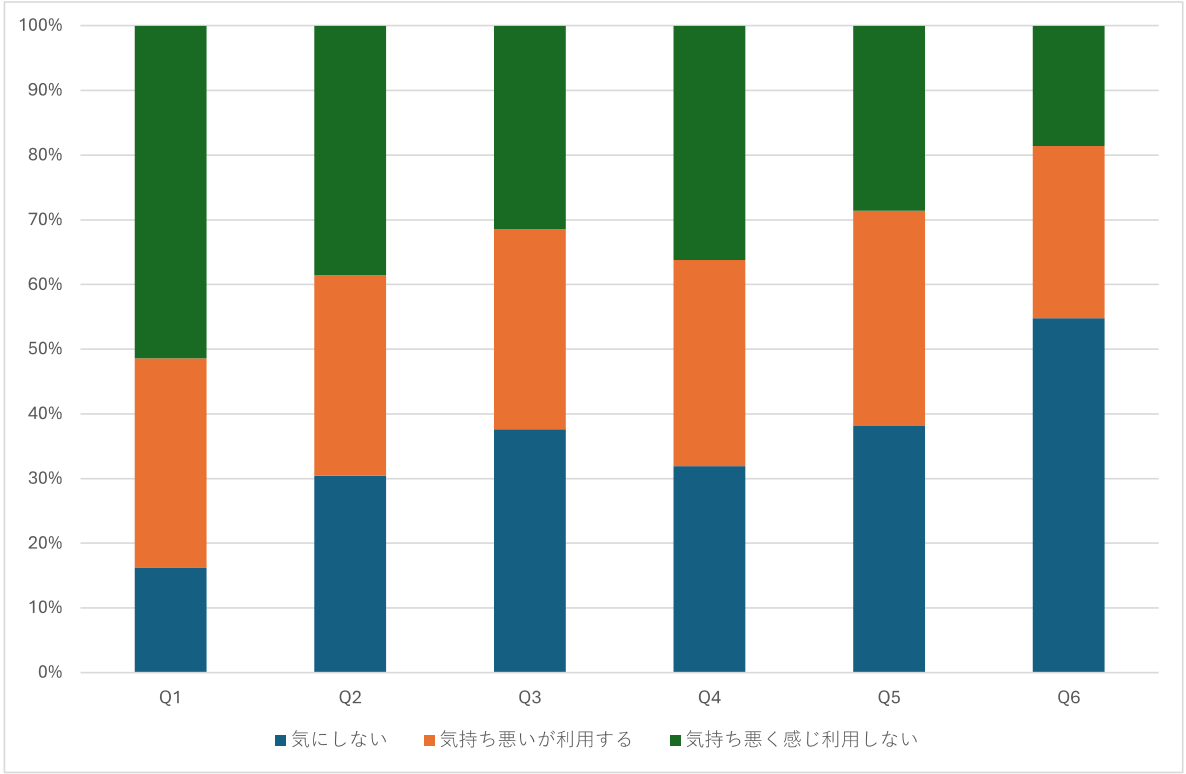
- 方法：クラウドソーシングでアンケート調査
- 参加者：210人
- アンケート内容
    1. 参加者属性について
    2. 架空のサービスに対しての印象
    3. 参加者のセキュリティ意識
    4. セッションリプレイサービスの認知

	属性	人数
年代	20代以下	18
	30代	84
	40代	71
	50代	30
	60代以上	7
性別	男性	98
	女性	112
ウェブサイト閲覧頻度	毎日	199
	週5~6日	10
	週3~4日	1
	週2以下	0

参加者属性

架空のサービスに対する許容度調査結果

	質問文
Q1	あなたのスクロール操作が全て記録されるウェブサービス
Q2	あなたのマウスの動きから悩んでいる商品を推測し，類似したおすすめの商品を推薦するショッピングサイト
Q3	あなたの閲覧時間が記録されているウェブサイト
Q4	あなたのキーボード入力内容から最適な広告を配信するウェブサイト
Q5	あなたがそのページを何回閲覧したかを記録しているウェブサイト
Q6	あなたのマウスの動きからページのデザインを改善しているウェブサイト



- 取得されることに対して否定的な傾向
- 情報の利用目的に応じて印象が変わる

許容度とユーザ属性の検定

ユーザの許容度とITスキル、セキュリティ意識の独立性について検定を行った

ITスキルでは全ての質問で有意差なし

セキュリティ意識ではQ1, 3, 5で有意差が見られた

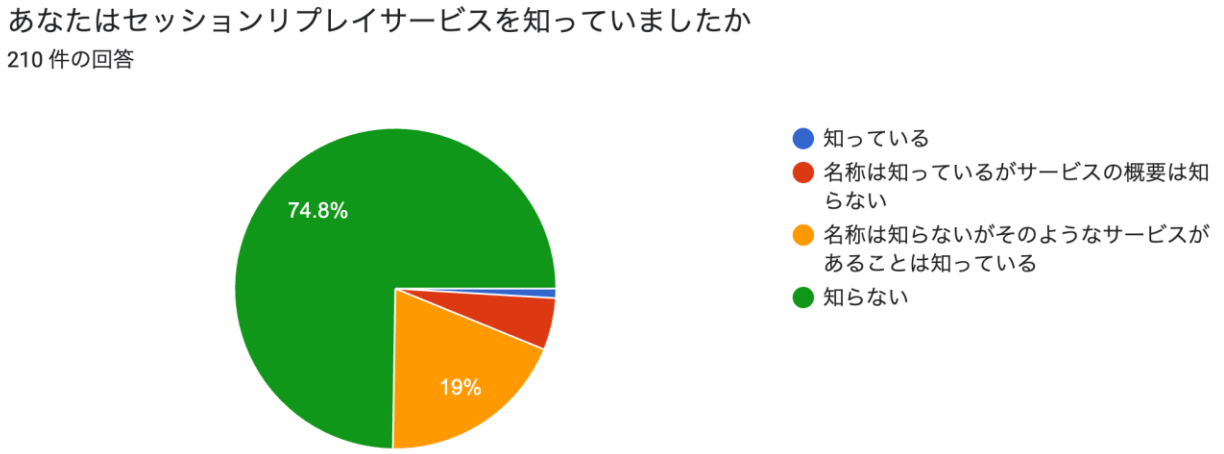
質問	ITスキル		統計量	*p*値
	あり	なし		
Q1	4	30	0.9579	0.3277
Q2	10	54	0.2522	0.6155
Q3	14	65	0.0009	0.9759
Q4	11	56	0.0978	0.7545
Q5	14	66	0.0013	0.9717
Q6	22	93	0.4001	0.5271

ITスキルとの検定結果

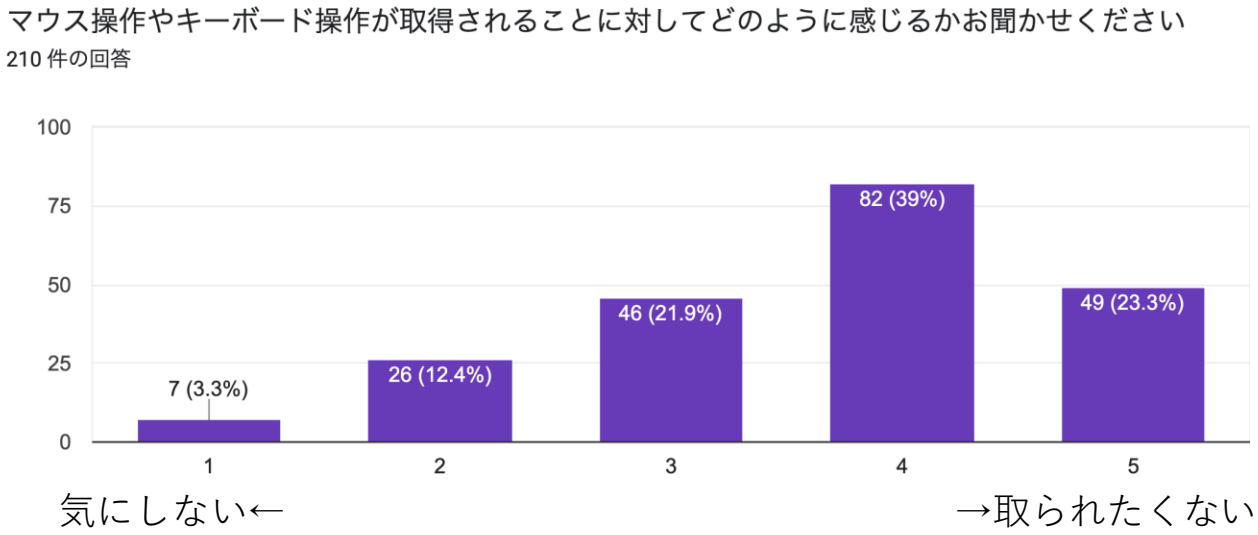
質問	セキュリティ意識		統計量	*p*値
	あり	なし		
Q1	27	7	7.5721	0.0059
Q2	43	21	3.1260	0.0771
Q3	56	23	8.5108	0.0035
Q4	41	26	0.3881	0.5333
Q5	57	23	9.1860	0.0024
Q6	70	45	0.8037	0.3700

セキュリティ意識との検定結果

セッションリプレイサービスの認知調査の結果



- 知っているユーザは2人(1%)
- **157人(74.8%)**がこのようなサービスがあることを知らない

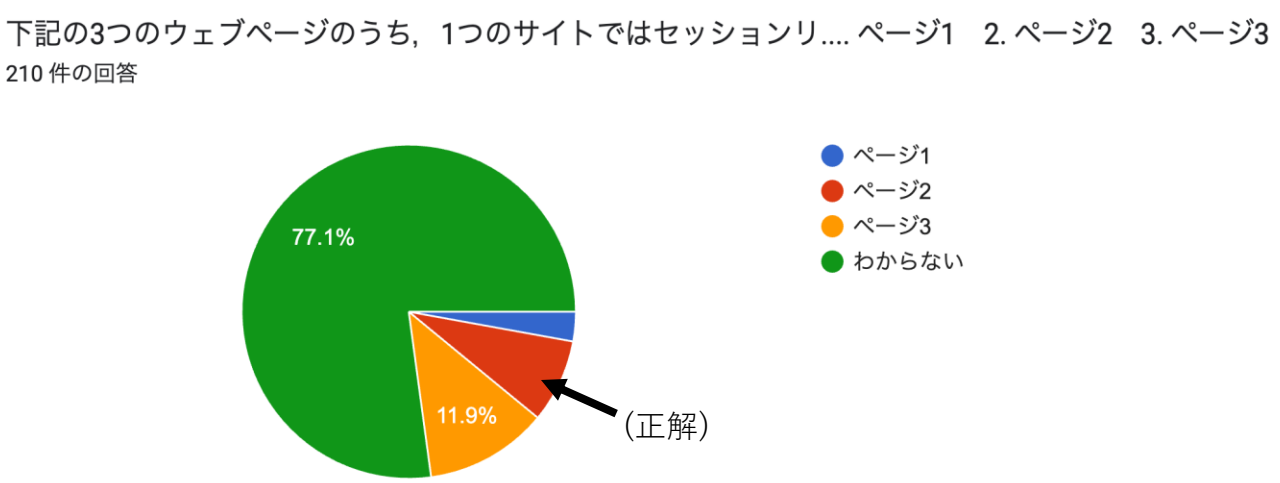


- **131人(62.3%)**が取得されることに対して否定的な回答

セッションリプレイサービスの認知度

ユーザがセッションリプレイサービスの利用に気づけるかを調査

1. 同一のウェブサイトを3つ用意
2. 内1つにマウスとスクロールの動きをサーバに送信するプログラムを設置（ページ2）
3. 参加者にアクセスしてもらいどのページでセッションリプレイサービスを利用していたかアンケート



17人(8.1%)が正解

→9割のユーザが**利用されていても気づかない**

ブラウザ上における個人識別性

本大学の学生15人のウェブサイト使用時のマウスとスクロール操作ログを取得

ランダムフォレストで2つの操作ログの操作者が同一か判定を行った

- 実験(2)-1

  「☆」描出操作時のマウスの動きを15人×5回取得

- 実験(2)-2

  ウェブサイトから質問の答えを探し出すタスクを3回依頼し

  マウスとスクロール操作を取得

実験1に使用したサイト

説明変数	実験1	実験2
マウス操作に要した時間	✓	-
カーソル速度の平均値，中央値，標準偏差	✓	✓
軌跡のx軸の最大値，最小値	✓	✓
軌跡のy軸の最大値，最小値	✓	✓
描出した軌跡のx軸の最大値，最小値	✓	-
描出した軌跡のy軸の最大値，最小値	✓	-
スクロール速度の平均値，中央値，標準偏差	-	✓
総移動距離	✓	-
データ取得数	✓	-

ブラウザ上の個人識別性　実験結果

	実験1	実験2
Accuracy	0.3697	0.3618
Precision	0.4333	0.2786
Recall	0.3502	0.3092
F_1	0.3407	0.2681

取得したログから100回識別を行い平均値を算出

再現率，適合率ともに0.27以上で識別可能

ランダムに識別を行う場合と比較して4倍の精度

導入状況とプライバシーポリシーの調査

導入状況

- 対象

  サイト：Trancoデータセットのうち**17,582サイト**

  サービス：14サービス

- 調査方法
  1. ブラウザ自動化ツールSeleniumでサイトにアクセス
  2. アクセス時に外部サーバに送信されるHTTPリクエストをデータベースに格納
  3. 既知のセッションリプレイサービスのURLからスクリプトを 読み込んでいるサイトを特定
  4. Norton Safe Webのカテゴリに従いサイトの業種を分類

プライバシーポリシーの調査

- 対象

  サービスごとにランダム50サイト

  合計400サイト

- 調査項目
  - 情報の利用目的の記載の有無
  - 具体的な利用サービス名
- 方法
  - 手動

国内外の導入状況

サービス	国内[1]	国外
Microsoft Clarity	702	818
Hotjar	89	1127
Mouseflow	68	86
Yandex	4	791
ContentSquare	20	88
Crazyegg	40	380
Dynatrace	2	39
foresee	1	35
fullstory	6	110
glassbox	2	10
inspectlet	0	31
logrocket	0	6
luckyorange	7	33
Smartlook	2	21
合計	943	3575

3575/17582(20.3%)

→国内(8.2%)より導入割合が多い

国内ではClarityが大部分を占める

国外ではHotjar, Yandexなど幅広く利用されている

導入サイトの業種カテゴリ

国内		国外	
カテゴリ	数	カテゴリ	数
ビジネス/経済	181	テクノロジ/インターネット	857
テクノロジ/インターネット	150	ビジネス/経済	481
買い物	130	教育	441
未分類	60	ニュース	258
健康	57	買い物	151
旅行	41	健康	134
娯楽	40	金融	128
教育	39	娯楽	85
ニュース	30	Web ホスティング	77
金融	28	未分類	76
仕事探し/キャリア	22	疑わしい	76
スポーツ/レクリエーション	21	政府/法務	73
車両	14	旅行	72
レストラン/食品	13	参考	66
社会/日常生活	11	ゲーム	59
不動産	10	オフィス/ビジネスアプリケーション	51

Norton Safe Webによる導入サイトの上位14カテゴリ

国内外ともに

- ビジネス/経済
- テクノロジ/インターネット

が上位を占めている

その他のカテゴリでも共通するものが多い

プライバシーポリシー調査結果

SRS01	国内[1]			国外		
記載	調査数	目的あり	サービス名あり	調査数	目的あり	サービス名あり
Clarity	50	47	12	50	47	3
Hotjar	50	47	6	50	49	6
mouseflow	50	31	2	50	47	8
crazyegg	50	41	10	50	50	6
contentsquare	27	24	2	50	49	4
luckyorange	20	14	0	27	27	0
Fullstory	5	4	1	50	48	6
Yandex	15	11	1	7	7	2
dynatrace	8	8	0	32	29	1
glassbox	4	4	0	8	7	0
smartlook	14	7	0	13	11	0
Foresee	2	2	0	0	0	0
inspectlet	3	3	0	7	7	0
logrocket	2	1	1	6	6	0
合計	300	244	37	400	384	36

目的

国内：**244**件(81.3%)

国外：**384**件(96%)

サービス名

国内：**37**件(12.3%)

国外：**36**件(9%)

目的は国内＜国外

サービス名は差がなく
9割が非公表

おわりに

ユーザ感情の調査

ユーザはマウスやキーボードの情報を取得されることに対して**否定的**

セッションリプレイサービスを利用されていても気づくことができない

操作ログからの個人識別性

約**30%**で識別可能

セッションリプレイサービス導入状況

国内：**8.2**%，国外：**20.3**%

導入サイトの傾向

国内外どちらもビジネス/経済、テクノロジ/インターネットなどのカテゴリが多い

その他のカテゴリでも共通するものが多い

プライバシーポリシーでの公表状況

目的の記載のあるサイトは国内：**81.3**%，国外：**96**%

サービス名の記載のあるサイトは国内：**12.3**%，国外：**9**%

今後の課題

ユーザに対して情報の取得をわかりやすく伝えるための研究開発