

# Webサイト上のターゲット広告の 観測システムの開発

---

明治大学総合数理学部 青山 綾佳○, 梶間 大地, 菊池 浩明

# 研究背景：トラッキングの透明性

2021年4月 Apple社

「App Tracking Transparency（アプリのトラッキングの透明性）」

他者のアプリやWebサイトを横断して  
ユーザの行動を追跡（**トラッキング**）  
する場合にユーザの許可を得ることが  
義務付けられた



# 研究目的

閲覧履歴による**ターゲット広告**の調査

RQ1 どの**ユーザ**がターゲットになりやすいか

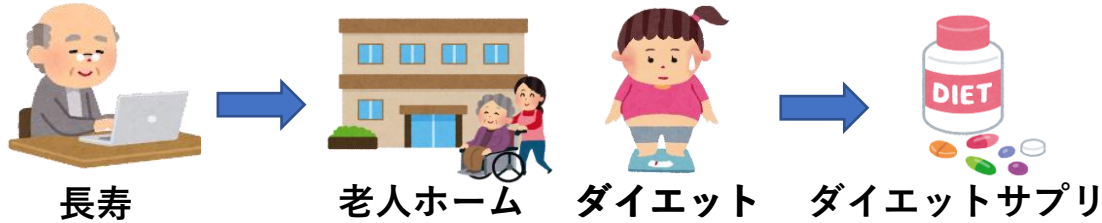
RQ2 どの**サイト**でターゲット広告が多いか



# 研究準備

## 興味の異なるユーザ(ペルソナ)

- ① Googleアカウントを9つ作成
- ② 興味に適した検索ワードを設定
- ③ 関連するサイトを閲覧



興味	検索ワード
(1)ファッション	コーデ, パンプス
(2)長寿	健康サプリ, 老人ホーム
(3)ダイエット	ダイエットサプリ, ダイエットグッズ
(4)ゲーム	ゲームソフト, ゲームアプリ

## 対象Webサイト

### (1) Q&A



### (3)観光



### (2) Web制作



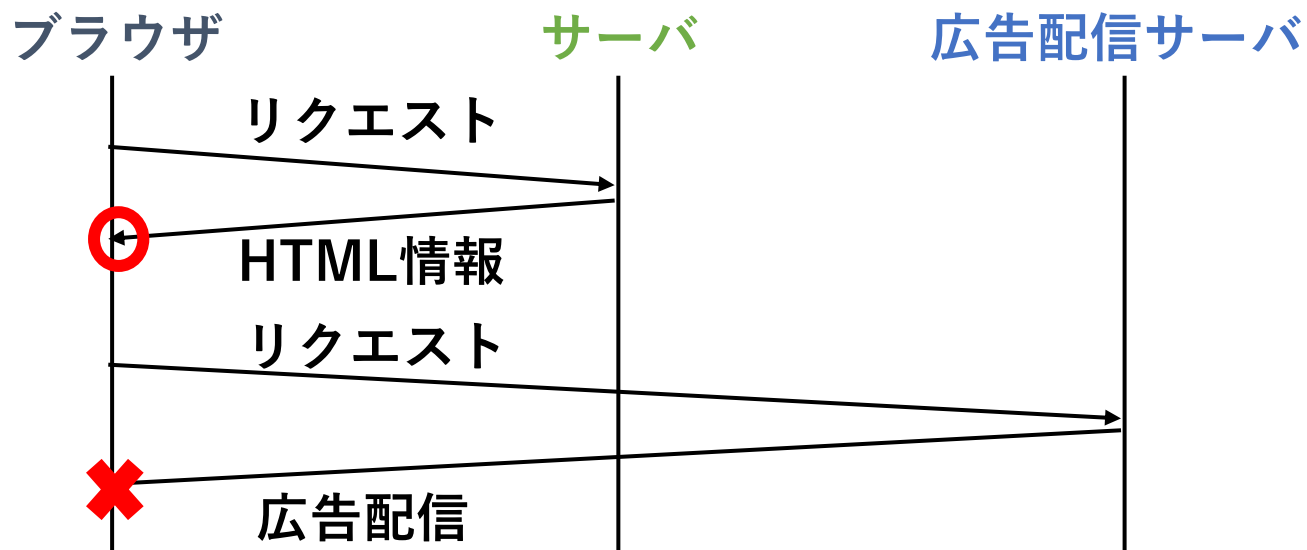
### (4) 家具



# ターゲット広告の観測の困難点

Web広告は閲覧時に入札され, どの広告が配信されるか決定する  
+  
ターゲット広告はユーザの情報をもとに配信される

通常のクローラー(Wget)では観測が難しい



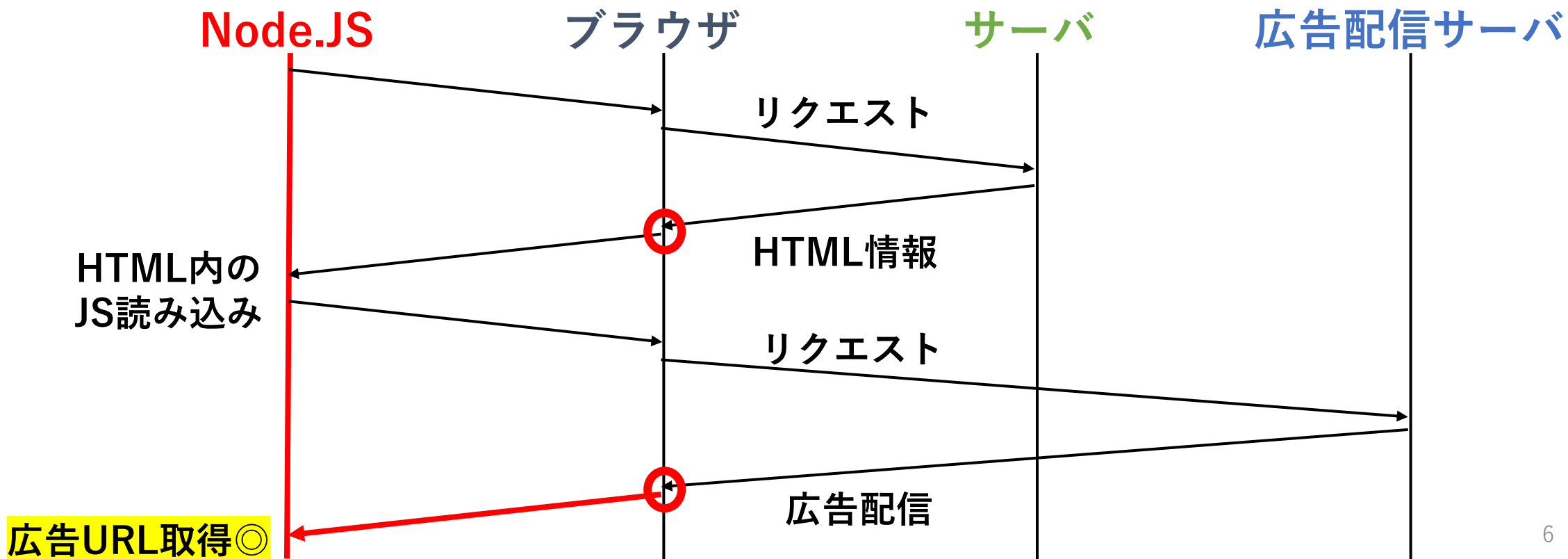
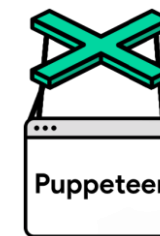
⇒ JavaScript実行後のサイトを取得できない



Wgetで表示したサイト  
(HTMLのみが読み込まれている)

# 本研究の解決法：Puppeteer

- Node.js 製ライブラリ
- サーバサイドである JavaScript の動作環境の Node.js 上で動作
- クローリング, キーボード入力やフォーム送信, クリックの自動化等の機能



# 実験の流れ（手動, 自動）

実験日時:2021年11月28日～12月6日

- ① ユーザごとの広告を調査するために設定したアカウントにログイン
- ② 対象Webサイトを検索し, そのサイトへ遷移する
- ③ サイト内の広告URLを取得

①



②



③



# 実験結果 (ターゲット広告数/全広告数)

システム

サイト概要 ペルソナ	Q & A	Web 制作	開発	ブログ	感想	漫画	攻略	お菓子	観光	家具	平均
育児	0/3	1/2	2/4	1/3	1/1	1/2	0/3	0/1	0/2	0/1	0.6/2.2
勉強	3/3	3/3	1/4	1/3	1/1	2/2	0/3	1/1	2/2	0/1	1.4/2.3
食事	2/3	3/3	2/3	3/3	1/1	2/2	2/3	1/1	0/1	3/3	1.9/2.3
ファッション	3/3	2/3	3/3	3/3	1/1	2/2	2/3	1/1	2/2	1/1	<b>2.0/2.2</b>
長寿	4/4	0/3	4/5	0/3	1/1	1/2	1/2	1/1	2/2	0/2	1.4/2.5
住宅	0/3	3/3	1/3	1/3	0/1	0/2	0/3	0/2	0/2	0/1	0.5/2.3
電子機器	4/4	1/3	1/3	2/2	1/1	0/2	0/3	0/2	3/3	0/1	1.2/2.4
ダイエット	0/3	0/2	3/5	0/3	0/1	0/2	0/3	0/2	0/2	0/1	0.3/2.4
ゲーム	0/3	0/3	0/4	0/3	0/1	0/2	0/3	0/2	0/2	0/3	0/2.6
平均 (自動)	1.8/3.2	1.4/2.8	<b>1.9/3.8</b>	1.2/2.9	0.7/1.0	0.9/2.0	0.6/2.9	0.4/1.4	1.0/2.0	0.4/1.6	

手動

サイト概要 ペルソナ	Q & A	Web 制作	開発	ブログ	感想	漫画	攻略	お菓子	観光	家具	平均
育児	4/4	3/3	2/7	6/7	3/3	6/7	4/5	0/6	8/14	1/3	3.7/5.9
勉強	5/5	3/3	2/7	2/5	2/3	1/7	5/5	1/6	4/13	1/3	2.6/5.7
食事	4/4	3/3	7/7	7/7	3/3	2/6	4/5	4/6	7/14	1/3	4.2/5.8
ファッション	3/4	3/3	7/7	5/5	2/3	6/7	5/5	2/5	6/14	1/3	4.0/5.6
長寿	4/4	3/3	8/8	3/5	3/3	7/7	5/5	1/6	11/14	1/3	<b>4.6/5.8</b>
住宅	4/4	3/3	3/7	0/7	1/3	7/7	5/5	4/4	5/13	3/3	3.5/5.6
電子機器	3/4	1/3	2/7	1/5	1/3	1/7	1/5	1/4	10/14	1/3	2.2/5.5
ダイエット	4/4	3/3	6/8	0/5	3/3	0/7	0/5	5/6	10/14	0/3	3.1/5.8
ゲーム	4/4	0/3	0/8	0/7	0/3	1/7	0/5	0/5	1/14	0/3	0.6/5.9
平均 (手動)	3.9/4.1	2.4/3.0	4.1/7.3	2.7/5.9	2.0/3.0	3.4/6.9	3.2/5.0	2.0/5.3	<b>6.9/13.8</b>	1.0/3.0	



# RQ1どのユーザがターゲットになりやすいか

サイト概要

感想	漫画	攻略	お菓子	観光	家具	平均
1/1	1/2	0/3	0/1	0/2	0/1	0.6/2.2
1/1	2/2	0/3	1/1	2/2	0/1	1.4/2.3
1/1	2/2	2/3	1/1	0/1	3/3	1.9/2.3
1/1	2/2	2/3	1/1	2/2	1/1	2.0/2.2
1/1	1/2	1/2	1/1	2/2	0/2	1.4/2.5
0/1	0/2	0/3	0/2	0/2	0/1	0.5/2.3
1/1	0/2	0/3	0/2	3/3	0/1	1.2/2.4
0/1	0/2	0/3	0/2	0/2	0/1	0.3/2.4

ターゲット広告が  
最も多いペルソナ  
手動：「長寿」 4.6

(システム：「ファッション」 2.0)

ターゲット広告が  
最も少ないペルソナ  
手動：「ゲーム」 0.6

(システム：「ゲーム」 0)

手動

ペルソナ	感想	漫画	攻略	お菓子	観光	家具	平均			
育児	4/4	3/3	2/7	6/7						
勉強	5/5	3/3	2/7	2/5						
食事	4/4	3/3	7/7	7/7						
ファッション	3/4	3/3	7/7	5/5						
長寿	4/4	3/3	8/8	3/5						
住宅	4/4	3/3	3/7	0/7						
電子機器	3/4	1/3	2/7	1/5						
ダイエット	4/4	3/3	6/8	0/5						
ゲーム	4/4	0/3	0/8	0/7						
平均 (手動)	3.9/4.1	2.4/3.0	4.1/7.3	2.7/5.9	2.0/3.0	3.4/6.9	3.2/5.0	2.0/5.3	6.9/13.8	1.0/3.0

# RQ2どのサイトでターゲット広告が多いか

ターゲット広告率が**高い**サイト  
手動：「Q&A」 95%

感想	漫画	攻略	お菓子	観光	家具	平均
1/1	1/2	0/3	0/1	0/2	0/1	0.6/2.2
1/1	2/2	0/3	1/1	2/2	0/1	1.4/2.3
1/1	2/2	2/3	1/1	0/1	3/3	1.9/2.3
1/1	2/2	2/3	1/1	2/2	1/1	<b>2.0/2.2</b>
1/1	1/2	1/2	1/1	2/2	0/2	1.4/2.5
0/1	0/2	0/3	0/2	0/2	0/1	0.5/2.3
1/1	0/2	0/3	0/2	3/3	0/1	1.2/2.4
0/1	0/2	0/3	0/2	0/2	0/1	0.3/2.4

ターゲット広告率が**低い**サイト  
手動：「家具」 33%

手動

ヘルプ										
育児	4/4	3/3	2/7	6/7						
勉強	5/5	3/3	2/7	2/5						
食事	4/4	3/3	7/7	7/7						
ファッション	3/4	3/3	7/7	5/5						
長寿	4/4	3/3	8/8	3/5						
住宅	4/4	3/3	3/7	0/7						
電子機器	3/4	1/3	2/7	1/5						
ダイエット	4/4	3/3	6/8	0/5						
ゲーム	4/4	0/3	0/8	0/7						
平均 (手動)	<b>3.9/4.1</b>	2.4/3.0	4.1/7.3	2.7/5.9	2.0/3.0	3.4/6.9	3.2/5.0	2.0/5.3	<b>6.9/13.8</b>	<b>1.0/3.0</b>

# システムと手動の実験結果

広告数が**多い**サイト  
 「観光」  
 結果が大きく異なる  
 手動：13.8  
 システム：2.0

感想	漫画	攻略	お菓子	観光	家具	平均
1/1	1/2	0/3	0/1	0/2	0/1	0.6/2.2
1/1	2/2	0/3	1/1	2/2	0/1	1.4/2.3
1/1	2/2	2/3	1/1	0/1	3/3	1.9/2.3
1/1	2/2	2/3	1/1	2/2	1/1	2.0/2.2
1/1	1/2	1/2	1/1	2/2	0/2	1.4/2.5
0/1	0/2	0/3	0/2	0/2	0/1	0.5/2.3
1/1	0/2	0/3	0/2	3/3	0/1	1.2/2.4
0/1	0/2	0/3	0/2	0/2	0/1	0.3/2.4

広告数が**少ない**サイト  
 「Web制作」  
 同じ結果を複数確認  
 手動：3.0  
 システム：2.8

手動

ヘルプ										
育児	4/4	3/3	2/7	6/7						
勉強	5/5	3/3	2/7	2/5						
食事	4/4	3/3	7/7	7/7						
ファッション	3/4	3/3	7/7	5/5						
長寿	4/4	3/3	8/8	3/5						
住宅	4/4	3/3	3/7	0/7						
電子機器	3/4	1/3	2/7	1/5						
ダイエット	4/4	3/3	6/8	0/5						
ゲーム	4/4	0/3	0/8	0/7						
平均 (手動)	3.9/4.1	2.4/3.0	4.1/7.3	2.7/5.9	2.0/3.0	3.4/6.9	3.2/5.0	2.0/5.3	6.9/13.8	1.0/3.0

# システムは手動よりも効率的

実験	1サイト当たりの観測時間	比
自動	28秒	1.0
手動	1分4秒	<b>2.3</b>

⇒ **2.3倍**の速さで実験可能, 手動より効率的である

# システムの誤差はユーザとサイトによる

誤差\サイト	Q&A	Web制作	開発	ブログ	感想	漫画	攻略	お菓子	観光	家具	平均
相対誤差(%)	22.0	6.7	48.0	50.8	66.7	71.0	42.0	73.6	85.5	46.7	51.3

$$\text{相対誤差} = \frac{|\text{手動で取得した広告数} - \text{自動で取得した広告数}|}{\text{手動で取得した広告数}}$$

⇒システムの誤差はユーザ及びサイトによって変動する

# 考察

- ・「観光」のサイトで誤差が大きくなった  
⇒サイト内の動画広告が多かった  
本システムでは動画広告を取得できないことが原因である
- ・広告URLを取得せずに処理を終えることがあった  
⇒実行中にシステムがタイムアウト  
システムによるページ遷移は回線やサーバの混雑状況によって動作に時間がかかる

# まとめ

RQ1 どのユーザーがターゲットになりやすいか

A. 長寿（健康サプリ, 老人ホーム）



RQ2 どのサイトでターゲット広告が多いか

A. Q&A（95%）



⇒ 更なる時間短縮や精度向上を今後の課題とする