

ウェブ検索における便益とプライバシーリスクを
考えさせる情報提示手法



ウェブ検索における便益とプライバシーリスクを 考えさせる情報提示手法

静岡大学情報学部
行動情報学科
清水 勇祐

Bachelor Thesis Series of Yamamoto Laboratory
Faculty of Informatics,
Shizuoka University
Copyright © 2021 Yusuke Shimizu

目次

| | |
|-----------------------------|----|
| 第1章　はじめに | 5 |
| 第2章　関連研究 | 7 |
| 2.1　ウェブ閲覧行動のトラッキング | 7 |
| 2.2　ウェブ検索ユーザのプライバシー意識 | 7 |
| 2.3　プライバシーリスクへの意識づけ | 8 |
| 第3章　提案手法 | 9 |
| 3.1　ふるまい | 9 |
| 3.2　第三者に知られる可能性のある閲覧履歴の検出方法 | 10 |
| 3.3　仮説 | 11 |
| 第4章　実験 | 12 |
| 4.1　実験協力者 | 12 |
| 4.2　実験条件 | 12 |
| 4.3　実験システム | 12 |
| 4.4　実験手順 | 13 |
| 第5章　結果 | 15 |
| 5.1　検索タスク時間 | 15 |
| 5.2　検索結果ページ閲覧時間 | 15 |
| 5.3　最大スクロール深度 | 16 |
| 5.4　履歴情報が表示された検索結果のクリック率 | 16 |
| 第6章　考察 | 18 |
| 6.1　検索行動への影響 | 18 |

| | | |
|---------|---------------|----|
| 6.2 | 仮説が支持されなかった原因 | 19 |
| 6.3 | 今後の課題 | 19 |
| 第7章 まとめ | | 21 |
| 参考文献 | | 22 |

図目次

| | | |
|-----|----------------------------|----|
| 1.1 | 提案システムを用いた際のウェブ検索結果画面の概観図. | 6 |
| 3.1 | 閲覧履歴のリーケの検出方法 | 10 |

表目次

| | | |
|-----|--|----|
| 4.1 | 検索タスクとシナリオ | 13 |
| 4.2 | プライバシー意識スコアの平均値（かっこ内は標準偏差） | 13 |
| 5.1 | 実験条件と検索クエリごとに見る、検索行動指標の平均値（かっこ内の 数字は標準偏差） | 17 |

第1章

はじめに

閲覧履歴に代表されるウェブ上の行動データは、ウェブ広告をより効果的に表示するために利用されている。行動データの収集はウェブ広告の効果を高めることができる。その一方で、過度なトラッキングが問題になっている。トラッキングの結果、過度に個人化された広告に対し不信感を抱く利用者や [4]、収集された自身の個人データの行方に不安を覚える利用者が存在する [2]。また、過度なトラッキングの結果、健康上の悩みや思想信条など、プライベートに踏み込んだ情報が反映された広告が表示される可能性も存在する。そのため、プライベートを把握される恥ずかしさに対する懸念が生じている [3]。

トラッキングを強めることでサービス提供者はより多くの広告収益を得ることができる。一方で利用者のプライバシーが犠牲になるため、サービス提供側の収益と利用者側のプライバシーのバランスが重要となる。しかし、どこまでのトラッキングを許容できるかは利用者によって異なるため、サービスの提供者や法律などが一律の基準を設けることは困難である。また、行動データが収集されていることを理解しているユーザは一定数存在するものの、自身の情報がいつ何によって紐付けられているのかを理解している利用者は少ない。結果的に、ウェブ検索中のユーザは自身のプライバシー保護のためにどのような行動を取るべきかを考える機会がないという問題が生じている。

本稿では、ウェブ検索中のユーザがウェブ閲覧時に生じるプライバシーリスクを意識し、自分が得られる便益との釣り合いを検討するための検索インターフェースを提案する。具体的には、検索結果リスト中に含まれるウェブページを閲覧することで閲覧先に漏れる可能性のあるページ閲覧履歴について、ウェブ検索結果画面に表示する。

図 □ に、提案システムの動作例を示す。提案システムは、一般的なウェブ検索エンジン上で動作する。ユーザが検索クエリを入力すると、提案システムは検索結果リストに表示された各ウェブページからサードパーティクッキーを検出する。同様に、ユーザが過去

<https://rentry.jp/note/web-camaera-osusume>

【2020】テレワークやビデオ会議におすすめのWEBカメラ10選と...

Rentryノートのメールマガジンに登録すると、新着記事のお知らせや、GoProなどカメラ・ガジェット・家電のお得な情報、最新家電が送料無料でレンタルできるRentryで使えるお得なクーポンがもらえます。メールマガに登録

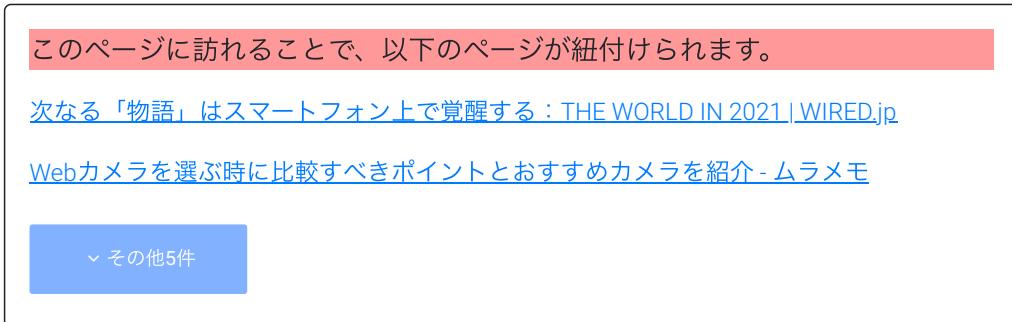


図 1.1 提案システムを用いた際のウェブ検索結果画面の概観図.

に訪れたページからも、ページを表示する際に読み込まれるサードパーティクッキーを取得する。それらクッキー情報を突き合わせることで、ウェブページを閲覧することで閲覧先に漏れる可能性のあるページ閲覧履歴をウェブ検索中に表示する。図 1-1 の例ではウェブカメラについて検索した場合に、検索結果とともに過去に閲覧したウェブページのタイトルが提示される。

ウェブページ閲覧における、プライバシーリスクの理解を支援する関連研究としては、機械学習を用いてプライバシーポリシーを解析しリスクを分類して表示するシステム [4] に関する研究が存在する。しかし、プライバシーリスクに敏感でないユーザにリスクを意識させるには、プライバシーリスクの種類を分類するだけでなく、リスクを実感できるような情報提示手法が必要となる。本稿で提案するシステムは、過去に閲覧したウェブページ名を具体的に提示することで、ウェブ検索中のユーザに漏れる可能性のある情報を分かりやすく実感させることが可能となる。

第2章

関連研究

2.1 ウェブ閲覧行動のトラッキング

Englehardt ら [5] はサードパーティクッキーによるトラッキングが、IP アドレスなどの既存の手段と比較してより効果的にユーザの属性を明らかにできることを明らかにした。Libert ら [6] は多くのウェブサイトにおける、サードパーティによるトラッキングの実態を調査した。ウェブサイトは収益を得る方法や、ホスティングなどの運営基盤に関してサードパーティーに依存する事が多く、結果的に不透明なトラッキングが行われていることを指摘している。上記の研究は、サードパーティクッキーの危険性を実証している。

ユーザのプライバシー保護を目的としたサードパーティクッキーをブロックする取り組みは数多く行われている。しかし、代替となる技術の存在により、ユーザーのプライバシー保護のための根本的な解決にはなっていない。Acar ら [7] はサードパーティクッキーの代替となる技術の出現による、ユーザのプライバシーへの悪影響を指摘している。

ユーザのプライバシー保護のためには、技術的な解決策では不十分であり、ユーザのプライバシー意識を向上させることが重要である。その支援を目的として、本研究ではユーザがトラッキングを把握した上で、プライバシー保護のための対応を考えることができるシステムを提案する。

2.2 ウェブ検索ユーザのプライバシー意識

Blase ら [8] は、サードパーティによるトラッキングに対するユーザの意識を調査した。その結果、ウェブ検索ユーザが、閲覧行動データを収集されても構わないと考える条件は、状況に依存することが多く、複雑であることを明らかにした。Leon ら [9] は、ウェブ

検索ユーザのデータ共有意欲に対し、広告会社が行っているプライバシーへの取り組みが与える影響を調査した。調査の結果、どのようなデータを誰が収集するのかを細かく制御できるのであれば、個人データを共有する意欲を高めることができることを発見した。以上の研究結果は、ウェブ検索ユーザは収集される情報に対する制御が可能であれば、情報が収集されることに懸念を感じないことを実証している。

また、Matic ら [9] は、高度に個人化された広告に対する被験者の反応を調査した。この研究では、高度な個人化を実現するために、被験者と親しい関係にある人物を広告の作成者としていた。被験者は、高度に個人化された広告に対して好意的な反応を示し、データの提供に値するものであると感じる傾向が見られた。上記の研究結果は、提供するデータの制御を維持した上での広告は、ユーザの利便性を向上させる可能性を示唆している。

本項ではウェブ検索ユーザが、収集される情報を自分で制御するための支援を目的とし、収集される情報をユーザ自身が選択できる検索インターフェースを提案する。

2.3 プライバシーリスクへの意識づけ

Harkous ら [4] は、ウェブページに訪れることで収集される可能性のある情報をサンキー図を用いて表現するというという手法を提案している。この研究では機械学習モデルを用いてプライバシーポリシーを分析し、収集される可能性のある情報を検出している。Kelley ら [10] は、食品の栄養ラベルを模した表示方法を用いて、ウェブサイトやウェブサービスが収集する情報をわかりやすく表現する手法を提案した。Harbach ら [11] は、スマートフォンアプリケーションにおけるプライバシーリスク意識の向上のための手法を提案した。提案手法では、アプリケーションのインストール時に権限を付与する画面において、許可される権限に応じた具体的な例を、ユーザが保存しているデータから取得し表示する。

本研究は上記の研究と同様に、プライバシーリスクへの意識づけを行う研究であると位置づけられる。しかし、能動的に使用しなければならない点で Harkous, Kelley らの研究と異なる。プライバシーリスクに敏感なユーザであれば、このようなシステムを能動的に利用できるため有効である。しかし、ウェブ検索ユーザの多くはプライバシーリスクに対して敏感ではない。このことから、ウェブ検索ユーザーに対し、プライバシー保護への具体的な対策を促すには不十分である。一方で、本研究はプライバシーリスクに敏感ではないユーザに対しても働きかけを行う点に関して、Harbach らと類似点が見られる。そこで本研究では、Harbach らを参考に、能動的でないユーザに対しても、プライバシーリスクへの意識づけを行うインターフェースを提案する。

第3章

提案手法

3.1 ふるまい

提案システムは、一般的なウェブ検索エンジンの拡張機能として動作する。ユーザの検索クエリを入力とし、以下の手順で、ウェブページを訪問すると第三者に知られる可能性がある閲覧履歴を同定し提示する。

1. 提案システムは、ウェブ検索結果に含まれる各ウェブページからサードパーティクッキーを取得する。
2. 同様に、ユーザのウェブ閲覧履歴からも、ページを表示する際に読み込まれたサードパーティクッキーを取得する。
3. 上記の手順によって得られたクッキーの発行ドメインを比較し、検索結果の各ウェブページに対して、発行ドメインが同一のクッキーを含む閲覧履歴を取得する。
4. 得られたウェブページの情報を「第三者に知られる可能性がある閲覧履歴」としてウェブ検索結果画面に表示する。

以上の提案システムによって、ウェブ検索ユーザは実際にウェブページを訪れる前に、第三者に知られる可能性のある閲覧履歴を知ることが可能となる。1章で述べた図①の検索結果画面を再び例にとり、以下に各手順の詳細を述べる。図①はユーザがビジネスメールの書き方について検索した場合の検索結果の一例を示している。この例では、2件の閲覧履歴が「第三者に知られる可能性のある閲覧情報」として表示される。提案インフェースによって、ユーザに対して以下のようなリスクについて考える機会を与えること

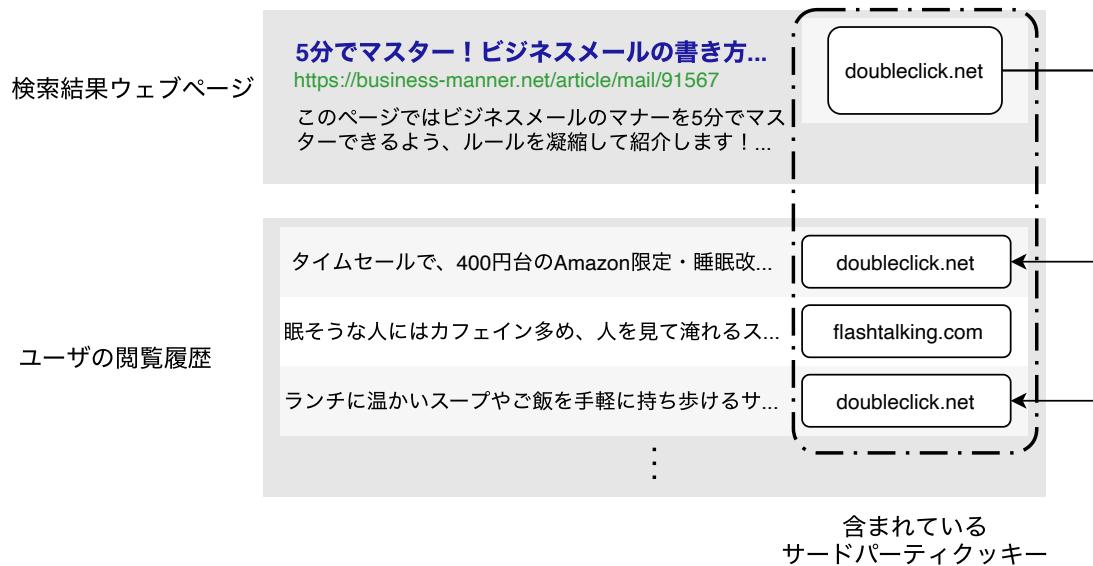


図 3.1 閲覧履歴のリークの検出方法

ができる。

- 表示された閲覧履歴に関する内容の広告が表示される可能性がある。
- 閲覧先のウェブページと紐付けられることで、自身の属性を分析されてしまうことにつながる。
- 閲覧しようとするウェブページも、この先他のウェブページと紐付けられる可能性がある。

3.2 第三者に知られる可能性のある閲覧履歴の検出方法

第三者に知られる可能性のある閲覧履歴の検出方法の概略を図 3.1 に記す。提案システムは、検索結果リストに含まれる各ウェブページのソースコードを解析し、発行されるサードパーティクッキーを取得する。この解析はオープンソースツールである webXray[12] を用いて行う。

次に、同様の手法を用いて、ユーザの閲覧履歴からも、アクセスした際に発行されたサードパーティクッキーを取得する。上記の手順によって得られたサードパーティクッキーの発行ドメインを比較し、検索結果の各ウェブページに対して、同一ドメインから発行されたサードパーティクッキーを持つ閲覧履歴を紐付ける。

3.3 仮説

本研究では、ユーザのウェブ検索におけるプライバシーリスクへの意識に対し、前述の提案システムが与える影響を調査する。具体的には、以下の仮説を検証することに焦点を絞る。

- H1 第三者に知られる可能性のある閲覧履歴を事前に表示することによって、ウェブ検索ユーザのプライバシーリスクに対する意識付けを高めることができる。
- H2 第三者に知られる可能性のある閲覧履歴を事前に表示することによって、ウェブ検索ユーザに対して、自身のプライバシー保護のために、どのような行動を取るべきか考える機会を与えることができる。
- H3 第三者に知られる可能性のある閲覧履歴を事前に表示することによって、ウェブ検索ユーザは表示される閲覧履歴が少ないページを優先して閲覧するようになる。
 - H3.1 データの共有に積極的になりやすいドメイン（ショッピングなど）では許容できる閲覧履歴の数が少なくなりにくい。
 - H3.2 データの共有に消極的になりやすいドメイン（健康問題、金融関連など）では許容できる閲覧履歴の数が少なくなりやすい。

第 4 章

実験

4.1 実験協力者

静岡大学の学部生、および大学院生から実験協力者を募集し、13名の実験協力者を集めた。実験協力者には、報酬として300円を支払った。

4.2 実験条件

実験協力者を提案インターフェースを使用する実験群と、使用しない統制群の2つに分けた。実験群が閲覧する検索結果一覧には、第三者に知られる可能性のある閲覧履歴を表示した。統制群には提案インターフェースを使用せず、第三者に知られる可能性のある閲覧履歴を表示しない検索結果ページを見て情報探索を行ってもらった。各実験条件への実験協力者の割り当てはランダムに行った。

4.3 実験システム

実験システムは、ウェブブラウザ拡張として実装したものを、実験協力者のコンピュータ上で動作させた。実験協力者が実験システム内で用いる検索システムは、Google検索などの一般的な検索エンジンを模したものを作成した。実験協力者が検索タスクを開始すると、検索クエリを入力した想定の検索結果画面が表示される。本実験では、検索トピックが検索行動に与える影響を調査するため、「おすすめのウェブカメラ」と「糖尿病の症状」という2種類のタスクを用意した。タスクの開始時には実験協力者に対し、表4ののようなシナリオを提示し説明を行った。本実験では、検索結果に表示するウェブページ

表 4.1 検索タスクとシナリオ

| 検索タスク | シナリオ |
|--------|---|
| ウェブカメラ | リモートでのやり取りが増えたので、ウェブカメラを購入したいと思ったとします。どれを購入するか迷ったあなたは、ひとまずウェブカメラについての情報収集のために、ウェブ検索することにしました。以下は、「ウェブカメラ おすすめ」というワードでウェブ検索した際の検索結果画面です。自分が購入したいウェブカメラが決まつたら検索を終了してください。 |
| 糖尿病 | 「頻繁に喉が渇く」「動悸がする」という症状が糖尿病の症状に当てはまるかどうか調べてください。以下は、「糖尿病 症状」とウェブ検索した際の検索結果画面です。症状に当てはまるかどうか調べられたと思ったら検索を終了してください。 |

表 4.2 プライバシー意識スコアの平均値（かっこ内は標準偏差）

| | 実験群 | 統制群 | p-value |
|-------------|----------------|----------------|---------|
| プライバシー意識スコア | 4.30 (0.42) | 4.36 (0.35) | 0.52 |

を、Google の検索エンジンにて、「ウェブカメラ おすすめ」「糖尿病 症状」と検索した際に表示される上位 100 件のウェブページから、提案手法によって収集される閲覧履歴が表示されるページと、そうでないページの数が同じになるように選択した。

4.4 実験手順

本実験は「事前アンケート」「タスク説明」「情報検索タスク」「事後アンケート」の 4 ステップから成る。

はじめに、実験協力者が本来持つプライバシー意識の差が、実験結果に影響を与える可能性を調査するため、タスクを行う前に実験協力者のプライバシー意識を計測した。アンケート項目には、菅沼らによる便益評価尺度 [13] から、プライバシー意識を測定するもののみを使用した。この尺度は 12 の質問項目から成り、実験協力者は 7 段階のリッカート尺度（1：全くそう思わない～7：非常にそう思う）で回答する。実験協力者のスコアは各回答の平均値で表される。したがってスコアは 1~7 の値をとる。分析には t 検定を行い、有意水準 5% で検定を行った。表 4.2 が示すように、実験群、統制群の間に統計的有意差は確認されなかった ($p = 0.52$)。このことから、2 群間にプライバシー意識の偏りによる差は存在しないといえる。

タスク説明のステップでは、タスクの大まかな流れの説明を行った。また、提案インターフェースを利用する実験協力者には、提案インターフェースにより、ウェブページ閲覧に伴い第三者に知られる可能性のある閲覧履歴が表示される旨を説明した。

情報検索ステップでは、実験協力者は前述の情報検索タスクに取り組んだ。実験協力者は検索結果リストに表示されたウェブページを閲覧し、質問への答えを考えた。検索結果がクリックされると、リンク先のウェブページがブラウザの新規タブに表示され、ウェブページの内容が閲覧可能となる。検索タスク中は、3章で述べた仮説の検証のため、以下の情報を収集した。

- 検索結果画面の閲覧時間
- 検索結果画面での最大閲覧深度
- 検索タスク中に閲覧したウェブページの件数
- 検索タスクの所要時間

提案インターフェースを使用した実験群に対しては、上記の情報に加えて、閲覧した各ウェブページに紐付けられると表示された閲覧履歴の件数も収集した。

実験協力者は、情報検索タスク終了後に、事前説明の際に回答したアンケートと同様のアンケートに回答した。上記アンケート、および事前アンケートから得られた回答を用いて、3章で述べた仮説を検証する。

第5章

結果

実験参加者 13 名から得られたデータのうち、事後アンケートに回答しなかった 1 名、および実験群の中で提案手法が表示されなかった 1 名を除いた 11 名（実験群：5 名、統制群：6 名）のデータを用いた。UI と検索トピックの 2 要因が、実験協力者の検索行動に与える影響について分散分析を行った。UI について、提案システム有り、無しの 2 水準を設定した。検索トピックについて、「ウェブカメラ」と「糖尿病」の 2 水準を設定した。なお、サンプル数が少ないため、正規分布に従うとの仮定のもと分析を行った。検定は有意水準 5% で行った。

5.1 検索タスク時間

実験協力者が検索結果をより慎重に閲覧したかどうかを調べるため、検索タスクにかけた時間を調査した。表 5.1 が示すように、実験群の平均タスク時間は、統制群の平均タスク時間と比較して、「ウェブカメラ」タスクで 53.84 秒、「糖尿病」タスクで 25.83 秒長かった。分散分析を行ったところ、実験条件要因および交互作用に統計的有意差は確認されなかった（実験条件： $p = 0.403$ 、交互作用： $p = 0.85$ ）。一方で、検索タスク要因に統計的有意差が確認された ($p < 0.05$)。この結果から、実験協力者は「ウェブカメラ」検索タスクに、「糖尿病」検索タスクよりも長い時間をかけたことが予想される。

5.2 検索結果ページ閲覧時間

実験協力者が検索結果をより慎重に閲覧したかどうかを調べるため、タスク中の検索結果ページ閲覧時間を調査した。表 5.1 が示すように、実験群の平均検索結果閲覧ページ時

間は統制群の平均検索結果ページ閲覧時間と比較して、「ウェブカメラ」タスクで 73.83 秒, 「糖尿病」タスクで 36.33 秒長かった。分散分析を行ったところ, 検索タスク, 実験条件, 交互作用いずれにも統計的有意差は確認されなかった（検索トピック : $p = 0.256$, 実験条件 : $p = 0.156$, 交互作用 : $p = 0.751$ ）。この結果から, 実験協力者が検索結果画面を閲覧する時間は, 提案インターフェースや, 検索トピックの影響を受けないことが予想される。

5.3 最大スクロール深度

実験協力者がどの程度検索結果を閲覧したのかについて調べるため, 実験協力者が実験中に閲覧したページスクロール量を調査した。表 5-1 が示すように, 実験群の平均スクロール深度は統制群の平均スクロール深度と比較して, 「ウェブカメラ」タスクで 3453.2 ピクセル, 「糖尿病」タスクで 2240.1 ピクセル長かった。分散分析を行ったところ, 検索タスク, 交互作用に統計的有意差は確認されなかった（検索トピック : $p = 3.127$, 交互作用 : $p = 0.540$ ）。一方で実験条件要因に統計的有意差が確認された ($p < 0.05$)。この結果から, 提案システムを用いることで, 実験協力者はより深い位置の検索結果を閲覧するようになったと予想される。

5.4 履歴情報が表示された検索結果のクリック率

検索トピックがプライバシー意識に与える影響を調べるため, 検索トピックごとに, 実験協力者が閲覧した検索結果のうち, 紐付けられる履歴情報が 1 件以上存在するページをクリックした確率（以下, クリック率）を調査した。表 5-2 が示すように, 「ウェブカメラ」タスクと比較して「糖尿病」タスクでは実験協力者のクリック率は低いことが伺える。t 検定を用いて分析を行ったところ, 検索クエリによるクリック率の違いに統計的有意差は見られなかった ($p = 0.22$)。この結果から, 検索トピックによって, 実験協力者が許容できる収集される情報の数に差は生じないと予想される。

表 5.1 実験条件と検索クエリごとに見る、検索行動指標の平均値（かっこ内の数字は標準偏差）

| 測定指標 | ウェブカメラ | | 糖尿病 | | p-value |
|----------------|----------------------|---------------------|----------------------|---------------------|-------------------------|
| | 実験群 | 統制群 | 実験群 | 統制群 | |
| タスク時間 (秒) | 632.17 (282.52) | 578.33 (340.49) | 359.50 (222.93) | 333.67 (136.83) | 0.025 0.403 0.85 |
| SERP 閲覧時間 (秒) | 311.83 (134.35) | 238.00 (158.46) | 219.50 (91.21) | 183.17 (100.73) | 0.256 0.156 0.751 |
| 最大スクロール深度 (px) | 5927.20 (4264.28) | 2474.00 (239.53) | 3721.60 (1986.20) | 1481.50 (859.29) | 0.127 0.009 0.540 |
| クリック率 | 0.36 (0.37) | 0.17 (0.25) | 0.17 (0.25) | 0.222 | |

第 6 章

考察

本章では実験で得られた結果の内容について考察する。本研究の目的はウェブ検索ユーザに対し、第三者に収集されている情報を可視化することで、自身のプライバシーを守るためにどのような行動を取るべきか考える機会を与えることである。

6.1 検索行動への影響

事前に立てた仮説 H1, H2 では、提案手法によってプライバシー意識を高め、自分のプライバシー保護のためにどのような行動を取るべきか考えさせることができると予想した。これを検証するため、タスク時間、検索結果ページ閲覧時間、最大スクロール深度を収集した。

実験群では提案手法により表示された情報により、収集されている情報が許容できるものかを考えよう促すことが期待できる。したがって実験群では、タスク時間、検索結果ページ閲覧時間いずれも長くなると予想した。しかし、表 5.1 が示すように、実験条件による統計的有意差は確認できなかった。このことから、提案インターフェースが影響を与えるとは言えない。すなわち、予想を支持する結果を得ることはできなかった。

また、実験群では提案手法により表示された情報により、慎重な検索行動を行うようになった結果、より深い位置の検索結果まで閲覧するようになると予想した。したがって実験群では、最大スクロール深度の値が大きくなると予想した。分析の結果、実験条件要因において、最大スクロール深度に統計的有意差が見られた。

以上のことから、タスク時間、検索結果ページ閲覧時間、およびスクロール深度の実験結果は仮説 H2 を支持したと考える。しかし、最大スクロール深度について、以下のような問題が存在する。ディスプレイの大きさなど、実験協力者の使用環境による変化の影響

が取り除けない。また、今回の実験では描画領域が肥大化しないよう、一定件数以上の履歴情報を折りたたんで表示した。しかし、実験協力者が展開した状態で検索タスクを続行した場合、その表示領域の分スクロール量が多くなってしまう。このことから、提案システムの効果を測定する指標として用いることは適切でない可能性が存在する。したがって「最上位から何件目の検索結果を閲覧したか」など、実験協力者の環境に影響されない指標を選択する必要があると考える。

事前に設定した仮説 H3.1、及び H3.2 では、検索トピックによって収集されることを許容できる閲覧履歴の件数が異なると予想した。これを検証するため、提案手法によって閲覧履歴が表示された検索結果のクリック率を収集した。「糖尿病」などの健康問題に関する検索トピックは、「ウェブカメラ」などの購買に関わるトピックと比較し、より自分の情報の提供に慎重になると考えた。したがってクリック確率が低くなると予想した。しかし、検索トピック要因による統計的有意差は確認できなかった。すなわち、実験結果は仮説 H3.1、及び H3.2 を支持しなかった。

6.2 仮説が支持されなかつた原因

今回の実験結果は仮説 H3 を支持しなかった。これに対し、2通りの原因が考えられる。第一に、実験協力者が自身の情報が収集されていることを知った上で、ウェブ検索によって得られる情報を優先した可能性である。その結果、統制群と類似する検索行動を行い、検索結果ページ閲覧時間やタスクの時間に影響が及ばなかったことが考えられる。第二に、実験協力者が収集される閲覧履歴が表示された場合に、内容に関わらず履歴が表示されているページを「危険」、表示されていないページを「安全」と単純に判断した可能性である。「安全」と判断したページを優先して閲覧し、十分な情報が得られなかつたと判断した場合にのみ「危険」と判断したページを閲覧するという行動を取つたと考えられる。その結果、統制群と比較し、閲覧するページの優先順位が変化したのみで、検索結果ページ閲覧時間やタスクの時間に影響が及ばなかつたことが考えられる。

6.3 今後の課題

実験で計測する指標について、3点の改善点があると考える。第一に、最大閲覧深度の計測方法である。6.1 節で述べたように、今回の実験で計測した最大スクロール深度では、実験協力者の使用するディスプレイなどの環境条件による影響を取り除くことができない。したがって、提案システムの効果を測定する指標として適切ではなかつた。検索結果

最上位から数えて何件目の結果を閲覧したかという情報など、実験協力者の使用環境に影響されない項目を計測する必要があると考える。

第二に、統制群の実験協力者が本来収集されていた履歴件数である。今回の実験では、統制群の実験協力者に対しては、閲覧履歴の分析を行わなかった。すなわち、今回の実験では実験協力者が閲覧履歴が収集されるページをどの程度クリックしたのかに対し、実験条件による影響を測定できなかった。統制群においても実験群と同様に閲覧履歴の分析を行い、表示される検索結果を閲覧することで収集される閲覧履歴を分析することで、履歴情報が収集される検索結果のクリック率を調査する。このクリック率を実験群のクリック率と比較することで、提案インターフェースが実験協力者自身のプライバシー保護のための行動を促すことができるか調査できた可能性が存在する。

第三に、プライバシーリスクを考えさせることができたかについて、事後のアンケート調査にて調査することである。今回の実験設計では、仮説 H1 について直接実証する指標を収集できていなかった。実験協力者の意識を調査するため、事後アンケートにて「閲覧するページを選択する際にプライバシーリスクについて考えたか」といった項目を設定し調査する必要があると考える。

今後の実験では、上記 3 件の情報を調査するように設計することで、提案システムの効果をより正確に分析できると考える。

第7章

まとめ

本稿では、ウェブ検索中のユーザのプライバシーリスクへの意識づけを支援するウェブ検索インターフェースを提案した。提案インターフェースは、ウェブ検索結果のリンク先と、ユーザが過去に閲覧したウェブページに含まれるサードパーティクッキーを比較し、第三者に知られる可能性のある閲覧履歴を表示する。提案インターフェースが、実際にウェブ検索ユーザのプライバシーリスクへの意識づけを高めることができるかを評価するため、ユーザ実験を行った。今回行った実験では、「ウェブカメラ」「糖尿病」の二種類の検索タスクを実施した。実験の結果、提案システムを用いた実験協力者は、より深い位置の検索結果まで閲覧する傾向が確認された。今後は十分な参加者を募集し、計測する情報などの実験設計を見直した上で再度実験を行い、提案手法を表示した際のユーザの検索行動の理解を深める必要がある。

参考文献

- [1] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, New York, NY, USA, 2012. Association for Computing Machinery.
- [2] Farah Chanchary and Sonia Chiasson. User perceptions of sharing, advertising, and tracking. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, p. 53–67, USA, 2015. USENIX Association.
- [3] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: Re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, New York, NY, USA, 2013. Association for Computing Machinery.
- [4] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *Proceedings of the 27th USENIX Conference on Security Symposium*, SEC'18, p. 531–548, USA, 2018. USENIX Association.
- [5] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. Cookies that give you away: The surveillance implications of web tracking. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, p. 289–299, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.
- [6] Timothy Libert and Reuben Binns. Good news for people who love bad news: Centralization, privacy, and transparency on us news sites. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, p. 155–164, New York, NY,

USA, 2019. Association for Computing Machinery.

- [7] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, p. 674–689, New York, NY, USA, 2014. Association for Computing Machinery.
- [8] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What matters to users? factors that affect users' willingness to share information with online advertisers. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, New York, NY, USA, 2013. Association for Computing Machinery.
- [9] Aleksandar Matic, Martin Pielot, and Nuria Oliver. "omg! how did it know that?": Reactions to highly-personalized ads. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, p. 41–46, New York, NY, USA, 2017. Association for Computing Machinery.
- [10] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, p. 1573–1582, New York, NY, USA, 2010. Association for Computing Machinery.
- [11] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 2647–2656, New York, NY, USA, 2014. Association for Computing Machinery.
- [12] Timothy Libert. Exposing the hidden web: An analysis of third-party HTTP requests on 1 million websites. *International Journal of Communication*, 2015.
- [13] 菅沼弥生, 西垣正勝, 大木哲史. 便益評価尺度を用いたプライバシー情報開示における合意形成プロセスに関する検討. コンピュータセキュリティシンポジウム 2019 論文集, Vol. 2019, pp. 756–763, Oct 2019.

謝辞

本研究の遂行ならびに論文の作成にあたり，ご指導を賜りました静岡大学情報学部講師山本祐輔先生に謹んで深謝の意を表します。本論文をまとめるにあたり，副査として有益な御助言と御教示を賜りました静岡大学情報学部准教授大木哲史先生に心より感謝の意を表します。本研究の遂行ならびに論文の作成にあたり御協力いただいた，静岡大学情報学部山本研究室の皆様に感謝致します。

2021年3月 清水勇祐