

Web ページ閲覧行動分析のためのブラウザ操作ログ収集ツール

Browser operation logger for analysing Web browsing behavior

杉田 賢治*1

Kenji Sugita

福原 知宏*2

Tomohiro Fukuhara

増田 英孝*1

Hidetaka Masuda

中川 裕志*3

Hiroshi Nakagawa

*1東京電機大学

Tokyo Denki University

*2産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

*3東京大学

The University of Tokyo

For assisting user's information seeking process on the Web, understanding his/her browsing behavior using a Web browser is important. We need to know various features of user's information seeking process; which keywords did s/he use, which point did s/he stop scrolling while s/he is looking at search results, and so on. In this paper, we describe a research tool called WeBOL which can record user's various Web browser operations. We also describe analysis results of a preliminary experiment using the WeBOL. We found two characteristic operations of users: (1) tab browsing operation, and (2) scrolling operation of a Web page.

1. はじめに

Web ページ内でユーザが注目した情報を判別し、情報の選択・閲覧の意図を探るためには、ユーザが Web ページ閲覧時にどのような行動を行っているかという特徴を知る必要がある。本研究ではユーザが閲覧した Web ページの中で有用な情報と不要な情報を識別するための基礎調査を目的とし、まずユーザが実際に Web ブラウザ上で行う操作を記録するためのツールを開発した。開発に当たってはユーザが Web ページを閲覧する際に Web ブラウザ上で行う操作に着目した。

ユーザが閲覧したページの履歴を用いた従来の手法では、ユーザの情報選択や閲覧に関する意図を探る上で不十分であり、Web ページ内でユーザが注目している情報を判別することも不可能である。それに対して本研究では、ユーザの Web 閲覧操作に関する詳細なデータを得ることにより、ユーザが閲覧した Web ページ内からユーザが注目した情報のみを抽出することが可能となる。その結果、ノイズの少ない Web ページを推薦することが可能になり、情報推薦の精度を向上させることが期待できる。

2. 先行研究

2.1 視線計測器の利用

高久らの研究 [1] では、Web ページ閲覧操作の分析に加え、眼球運動データの分析を行っている。ユーザの情報探索行動を記録する方法としては画面キャプチャ映像と視線計測データを利用しており、それらを人手で解析し、予め定義されたカテゴリに操作を当てはめていく。この研究では、実験により記録したデータの分類・解析を人手で行う必要があるため、結果を導くまでに時間が掛るといった問題が存在する。また、視線計測器を用いてユーザの視線を追跡する手法は、実験を行う環境に限られ、ユーザが日常的に Web ページを閲覧している状態と同じような環境で計測を行うことは困難である。

2.2 プロキシサーバの利用

Atterer らの研究 [2] では、Web ブラウザ上でのユーザの閲覧操作を記録するために、専用のプロキシサーバを介して Web

連絡先: 杉田 賢治, 東京電機大学院 未来科学研究科 情報メディア学専攻 修士課程, 東京都千代田区神田錦町 2-2, sugita@cdl.im.dendai.ac.jp

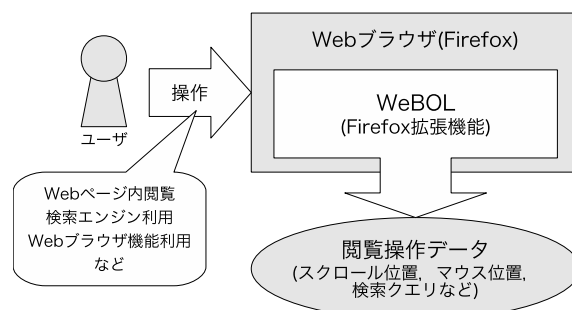


図 1: 本ツールの概要

ページにログ記録用のスクリプトを追加し、プロキシ経由での記録を行っている。この手法では Web ページ上の閲覧操作は取得できるが、タブ操作やツールバーの操作など、Web ブラウザの機能の利用についての記録は不可能である。

2.3 Web 操作閲覧記録ツールの要件

本研究で開発を行う Web 操作閲覧記録ツールは以下の 2 つの要件を満たす必要がある。

1. ユーザの Web ページ閲覧中の行動に関するデータの収集
2. 特別な機器を用意する必要が無く、日常的な Web ページ閲覧環境に近い取り扱いの容易なツール

1 番目の要件を満たすには、Web ページ内を閲覧するための Web ページ上での操作だけではなく、タブ等の Web ブラウザの機能を利用した Web ブラウザでの操作も自動的に記録する必要がある。そのためには Web ブラウザから直接情報を記録する仕組みを構築する必要がある。

2 番目の要件に関しては、より多くの被験者からデータを収集し、日常的な Web ページ閲覧環境に近づける必要があるためである。これらを満たすためには、被験者実験を行う上で容易に実験環境を構築できるようなツールが必要である。

3. WeBOL: Web 操作閲覧記録ツール

本研究で開発している Web 操作閲覧記録ツール WeBOL(Web Browser Operation Logger) は、ユーザが Web

ページを閲覧する上でユーザがどのような行動を行っているかを分析するため、ページ遷移を行うための各種操作、スクロール位置、マウスカーソル位置などの、Web ブラウザ上でユーザがページを閲覧する際に行う主要な操作を記録することを目的として開発を行っている。図 1 で示すように、WeBOL が組み込まれた Web ブラウザ (Firefox) をユーザが操作することにより、Web ページ閲覧時に閲覧操作に関するデータを取得することが可能である。

本研究では、ユーザが容易に扱えるツールを実現するため、Mozilla Foundation が提供する Web ブラウザ Firefox の拡張機能として、これによりユーザは本ツールを Firefox にインストールするだけで Web 閲覧時の操作が記録可能となり、多くの環境で動作することを可能にした。記録は全て Firefox 上の SQLite データベースに記録され、後に簡単にデータを抽出することが可能である。

4. 記録対象とする操作と分類

WeBOL が記録するユーザによる操作の一覧を表 1 に示し、それらの操作をどのような目的で記録するかの分類を行った。

表 1: Web ページ閲覧操作カテゴリ

操作名	定義
click	ページ内の任意の部分をクリックした
scroll	ページ内でスクロールを行った
mousemove	ページ内でマウスカーソルを動かした
selectText	ページ内のテキストをマウスで選択した
clickLink	リンクをクリックした
doSearch	検索エンジンで検索を行った
back	ブラウザの“戻る”機能を利用した
forward	ブラウザの“進む”機能を利用した
tabOpen	タブを新たに開いた
tabClose	タブを閉じた
tabSelect	タブを切り替えた
searchbarFocus	ブラウザ上部の検索バーにフォーカスした
searchbarSearch	ブラウザ上部の検索バーから検索した
urlbarFocus	ブラウザ上部の URL バーにフォーカスした
urlbarCommand	ブラウザ上部の URL バーの文字列を実行した

ページ内での操作

これらは現在閲覧している Web ページ内において、どの部分の情報に注目しているかの判別を行うための情報として利用することが期待される。特にスクロールの操作は、ユーザが Web ページ内のどの部分を長時間閲覧していたかという情報や、スクロールの速度を記録するためのデータとして利用することが狙いである。対象とする操作は表 1 の click, scroll, mousemove, selectText である。

新たなページ遷移の操作

新たなページへ遷移する場合、ユーザは既存のページ内のリンクから遷移する。この操作は、Web ページ内のどのリンクを選択したかという情報を取得するために利用する。対象とする操作は表 1 の clickLink である。

ページ間遷移の操作

既に閲覧した Web ページ間を遷移しているかを判別するために利用する。既に閲覧したページを再度閲覧したという情報として利用することが目的である。対象とする操作は表 1 の back, forward である。

検索エンジンの利用

検索エンジンから情報を検索している途中で、新たに検索を開始する場合や、現在の検索クエリを変更する場合が存在する。これらの操作はユーザの要求する情報の変化を理解するために利用することが期待される。対象とする操作は表 1 の doSearch である。

タブ操作

Web ブラウザのタブ機能を利用し複数の Web ページを並行して閲覧している場合、現在どのページを閲覧しているかを判別する必要がある。特に、新たなページに遷移する際に既存のタブではなく新たなタブとしてページを開いた場合は、そのページは有用である可能性が高いといった判断に利用することが狙いである。対象とする操作は表 1 の tabOpen, tabClose, tabSelect である。

ブラウザの標準機能の利用

これまでに挙げた操作の分類に当てはまらない操作の中から、Web ブラウザの機能をどの程度利用しているかを確認するための操作を記録する。ユーザによって機能の利用に差が出るかといった情報として利用することが目的である。対象とする操作は表 1 の searchbarFocus, searchbarSearch, urlbarFocus, urlbarCommand である。

5. 予備実験

事前に設定した仮説が有効かどうかを検証するために、WeBOL を用いてユーザが Web ページ閲覧時に Web ブラウザで行う操作を記録した。今回は 3 名の被験者に WeBOL を導入した Firefox 上で情報検索を行ってもらった。

5.1 実験前の準備

検索エンジンを利用して調べるという前提で、具体的な検索内容を被験者に自由に設定してもらい検索課題とした。また、実験の流れについても説明した。

5.2 実験開始から終了まで

実験を開始する際には、被験者にツールバー上の“記録開始”ボタンを押してもらい、必要な情報が満足に得られた時点でツールバー上の“記録停止”ボタンを押してもらうように指示した。

5.3 実験後のインタビュー

実験終了後、記録された履歴と画面映像等を参照しながら被験者にインタビューを行った。インタビューでは、各ページに対して、有用な情報が含まれていたか、そのページを選択した意図などを被験者に質問した。

6. 実験結果

表 2: 被験者ごとの記録データ数

	被験者 A	被験者 B	被験者 C
閲覧ページ数	22	33	33
スクロール回数	429	662	814
タブ操作	7	16	35
マウス移動	16650	9793	19122
検索バー利用	0	5	1

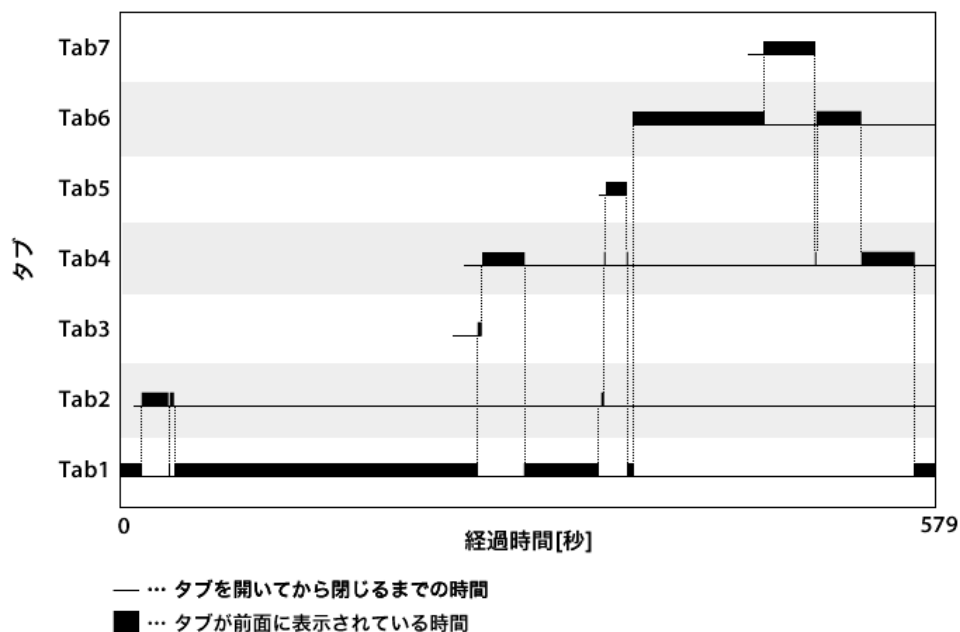


図 2: 被験者 C のタブ操作

Firefox 内部データベースに記録された Web 閲覧操作データを基に、被験者ごとの記録データ数を表 2 にまとめた。表 2 中の値を被験者ごとに比較した際、被験者ごとの特徴と被験者に共通する特徴が見られた。それぞれを以下に述べる。

6.1 被験者ごとの特徴

タブ操作では被験者 A と C を比較した場合、操作数に大きな違いが見られた。被験者 A は殆どタブに関する操作を行っておらず、新たに開いたタブは 2 つ程度であった。それに対し、被験者 C はタブ操作を頻繁に行っており、Web ページの内容を比較するために複数のタブを切り替えて利用していたことが実験中の画面キャプチャから確認できた。

マウス操作では被験者 B は他の被験者と比較した場合、被験者 B はデータ数が少ないことが分かる。このことから被験者 B はあまりマウスを動かす操作は行っていないことが分かる。実験中の操作を画面キャプチャで比較した際、被験者 A と C が Web ページ内で頻繁にマウスの移動を行っているのに対し、被験者 B は Web ページの内容を閲覧している際にはリンクの選択操作以外では殆どマウスの移動は行っていなかった。また、被験者の中には Web ページ内のテキストを読んでいる際に現在目で追っている文字列部分にマウスカーソルを移動する操作も見られた。このようなユーザ毎の特徴もマウス移動のデータ量に被験者間で差が出ていると理由だと考えられる。

6.2 被験者に共通する特徴

今回の予備実験では被験者に共通する特徴もみられた。特にタブ操作とスクロール操作が該当する。

タブ操作

タブ操作に関しては、新たに開いたタブは後に別のタブとの閲覧を往復する操作が見られた。また、一定時間残されたタブは後に何度も参照する場面も見られた。被験者実験後のインタビューでリンクを新たなタブで開いた理由に関して質問したところ、その Web ページに有用な情報が掲載されており、後に参照するために行ったという回答が得られた。

実際のタブの利用状況の例を被験者 C のデータを用いて図 2 に示す。横軸は経過時間、縦軸は開いたそれぞれのタブを表す。Tab1 は Web ブラウザの初期状態で開かれているタブであり、Tab2 以降は被験者が任意に開いたタブである。

この図から読み取れる事柄は、タブで開いたページの中にはしばらくの間閉じずに何度も参照している点である。これは図 2 中の Tab2, Tab4, Tab6 が該当する。また、タブを新たに開いた際にそのタブを操作していない場合が見られる。これはタブをバックグラウンドで開いたためである。これより新たに開いたタブは必ずしもすぐに閲覧するとは限らないことが分かる。これは Tab1, Tab6 以外の全てのタブにおいて見られる。

スクロール操作

スクロール操作に関しては、スクロール速度と着目している場所に一定の関係がみられた。スクロール操作のデータを比較するために、実験後のインタビューで被験者が注目した Web ページと注目しなかったと回答した Web ページを例にとり、双方の Web ページでのマウス操作を図 3 と図 4 に示した。横軸が Web ページを表示してから経過時間であり、縦軸が Web ページの縦のサイズを最大としたスクロール位置である。

注目した Web ページでは、スクロールを Web ページの一番下まで行っており、Web ページの内容をひと通り閲覧したことが分かる。また、グラフの傾きが比較的緩やかな箇所が多く存在し、これはスクロール位置でゆっくりとスクロールを行っていることを示している。図 3 のグラフからはこのような箇所が多く見られる。

図 4 の注目しなかった Web ページでは、図 3 のグラフから見られた緩やかな傾きは殆ど存在しない。この結果から、注目しなかった Web ページでは閲覧時間が少なく、スクロールをページ上部分でしか行わず、スクロールの速度も速いという特徴が見られる場合があるといえる。

スクロール操作に関してのその他の特徴としては、ユーザが Web ページ内で特定の情報を探している場合、上下スクロールを繰り返す傾向が見られた。

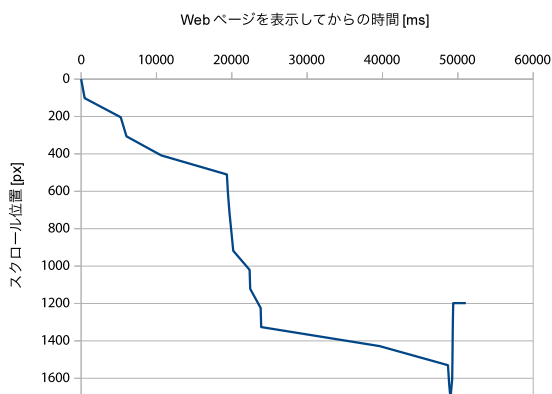


図 3: 注目した Web ページでの操作例

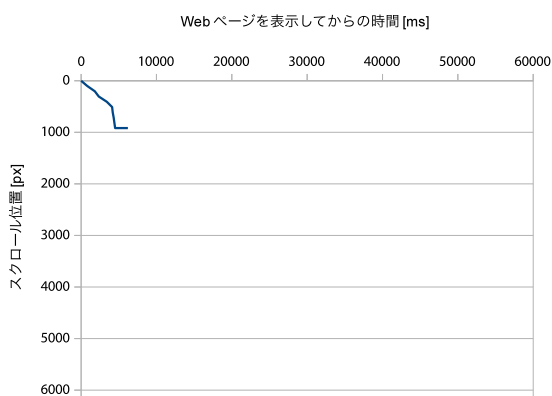


図 4: 注目しなかった Web ページでの操作例

7. 考察

7.1 被験者ごとの特徴に関する考察

実験結果より、被験者毎に操作の回数が大きく異なる部分が見られた。これは普段 Web ページを閲覧する際に行う操作の個々の特徴だと考えられる。このように個人毎に異なる特徴が確認できたことから、Web 閲覧時の操作するデータを情報推薦に応用する際には、ユーザ毎にプロフィールを生成する必要があると考えられる。

7.2 被験者に共通する特徴に関する考察

予備実験では被験者に共通する特徴がタブ操作とスクロール操作に見られた。これらに着目することでどのような特徴データとして利用可能かを考察する。

タブ操作に着目する場合、タブで複数の Web ページを閲覧しているという情報を得ることで、従来の手法では考慮されていなかった以下の特徴が識別可能になると考えられる。

- タブで開いた Web ページに関しては他の Web ページよりも有用な情報が含まれている
- またはタブで開いて一定時間経過した Web ページに関しては、ユーザが不要と考えた Web ページではない

被験者実験の結果からも分かるように、全ての被験者がタブ機能を利用していたので、Web 閲覧操作に着目する際はタブ操作を考慮することは重要であると考えられる。

スクロール操作に着目する場合、スクロールの位置と速度の情報を基に Web ページ上でユーザが注目した部分を識別することが可能になると考えられる。これにより、Web ページ単位の注目度ではなく、Web ページ内の領域ごとの注目度を識別するための特徴として利用可能になると考えられる。領域ごとの注目度が識別可能になれば、よりノイズの少ない特徴抽出が可能となると考えられる。

被験者実験の中では、Web ページ上で特定の情報を探している場合に上下にスクロールを繰り返す操作を行なっていることが確認できたが、同じく特徴的なスクロール操作に着目することで、ユーザの操作の意図を識別することが新たに可能になると考えられる。

以上のような被験者に共通する特徴から、以下の仮説は有効であるといえる。

- 後で閲覧したいページはタブ機能でページを開く
- スクロールの位置、速度から注目箇所が分かる

これにより、共通する操作を特徴データとして利用可能であるといえる。

8. おわりに

本研究では、Web ブラウジング時の操作に着目した情報推薦に関する研究を行うにあたり、ユーザの Web 閲覧操作を記録するための支援を行うツール WeBOL を開発した。これにより、Web ブラウザ上でのユーザの Web 閲覧操作を記録することが可能になった。

実際に WeBOL を利用するために、いくつかの仮説を立て、それらの仮説が有効かどうかの検証を行うための被験者実験を行い、WeBOL を用いてユーザが Web ページ閲覧時に Web ブラウザで行う操作を記録した。

その後、被験者実験より得られた Web 閲覧操作データからいくつかの特徴を発見し、いくつかの仮説が有効であると確認出来た。これにより本研究では、ユーザに共通する操作を特徴データとして利用可能であることが確認できたといえる。

本研究では予備実験より得られたデータを基にユーザの特徴を抽出したが、それらには未だ分析を行う余地がある。本論文で挙げた特徴以外にも新たな特徴を発見できる可能性を考え、今後更に詳細な分析を進める必要がある。

参考文献

- [1] 高久雅生, 江草由佳, 寺井仁, 齋藤ひとみ, 三輪真木子, 神門典子, “タスク種別とユーザ特性の違いが Web 情報探索行動に与える影響: 眼球運動データおよび閲覧行動ログを用いた分析”, 情報知識学会誌 早期公開 2010, Vol. 20, No. 3 pp.249-276 (2010).
- [2] Atterer, R., Wnuk, M., and Schmidt, A. “Knowing the User’s Every Move - User Activity Tracking for Website Usability Evaluation and Implicit Interaction”, Proceedings of the 15th international conference on World Wide Web (WWW ’06), pp.203-212(2006).