

情報の精査傾向を内省させるためのウェブ検索インタフェース

鈴木 雅貴[†] 山本 祐輔[†]

[†] 静岡大学大学院総合科学技術研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†]suzuki@design.inf.shizuoka.ac.jp, ^{††}yamamoto@inf.shizuoka.ac.jp

あらまし 本稿では、ユーザがウェブ検索を通じて健康トピックについて学習する際、ユーザの情報精査を促すインタフェースを提案する。提案インタフェースは、ユーザが閲覧したウェブページの品質を内容と見た目の観点から可視化し、ページ閲覧傾向を認知させることでユーザ自身のウェブ検索行動の内省を促し、見た目ではなく内容の品質の高いウェブページにアクセスしてもらうことを狙う。提案インタフェースの効果を検証するため、クラウドソーシングを用いてオンラインユーザ実験を実施した。ユーザ実験の結果、非大卒経験者が提案インタフェースを使用した場合、使用しない場合に比べて、情報探索の際にウェブページの内容の品質を重視し、内容の品質が高いウェブページを多く閲覧する傾向にあることが明らかになった。一方、大卒経験者が提案インタフェースを使用した場合、使用しない場合に比べて、内容の品質が高いウェブページを閲覧しない傾向にあった。

キーワード 情報信憑性、情報検索、ヒューマンファクター、ゲーミフィケーション

1 はじめに

近年、ウェブ検索を通じて学習することが一般的になりつつある。ウェブ検索を通じて情報を得ることは便利であるが、ウェブ情報は誤った情報が多いという問題がある。Sillence らによると、ウェブ上に存在する健康情報の約半分は専門家のチェックが入っていない品質の低い情報であることが報告されている [1]。また、ウェブ検索ユーザはデザインがきれいなウェブページを信頼できるサイトであると考える傾向 [2] や、分かりやすいウェブサイトを優先的に選んでしまう傾向 [3] があることが明らかになっている。このように、ウェブには品質の低い情報が多いにもかかわらず、ユーザが見た目に関する要素のみでウェブ情報を選択すると、誤った情報を選んでしまう可能性が高い。それゆえ、ウェブ検索ユーザにウェブ情報の見た目ではなく、データや証拠の有無といった内容を重視するように促すための情報アクセスシステムが必要である。

ウェブ検索ユーザの情報精査を促すための支援を行う研究はいくつか行われている。例えば、Leong らは、ある事実に対する証拠となる情報を提示する検索システムを作成した [4]。また、Ennals らは、ウェブ情報に反証が含まれている場合、ハイライトをすることで注意深い情報探索を促した [5]。こうしたシステムは、批判的に情報探索をしようとする態度を有するユーザには有効であると考えられる。しかし、上記で述べたように、ウェブ情報を批判的に精査するユーザは少ないため、これらのシステムが有効にはたらく場面は限られる。また、上記のシステムはユーザ自身の情報探索における傾向を提示していない。Kusumi らによると、批判的であろうとする態度には自分自身の行動を内省し、熟考することが重要であるとされている [6]。それゆえ、本稿ではウェブ検索ユーザの情報精査を支援するために、ユーザの情報探索における傾向を可視化し、内省を促すことを目指す。



図1 ReflectionUI の図。SERP の右側に内容の品質に関するスコアを可視化したゲージを提示している。

本稿では、ウェブ検索ユーザに内容の品質の高いウェブページを積極的に閲覧し、情報の精査を促すインタフェースを提案する。本稿では、上記のアイデアを実現するため、**ReflectionUI** と **SimpleUI** の2つのインタフェースを設計する。**ReflectionUI** は、図1に示すように、検索結果一覧ページ (SERP) の横に、ユーザの情報精査傾向を数値で表したゲージを提示する。**ReflectionUI** を使用したユーザが、内容の品質が高いページを閲覧した場合はゲージのスコアが上昇し、内容の品質が低いページを閲覧した場合はスコアが下降する。ユーザにはスコアの変動要因を明らかにしはしないが、情報探索過程で UI 上のスコアの変化と自身のページ閲覧行動の関係を内省させることで、内容の品質を重視した情報閲覧の重要性に気づかせることを狙う。ユーザは自身の情報精査傾向を内省し、スコアを上げるために内容の品質が高いウェブページを閲覧しようと努力することを期待する。

SimpleUI は、図2に示すように、検索結果一覧ページ (SERP) の横に、見た目と内容の品質のスコアを可視化した2つのゲージを提示している。**SimpleUI** を使用したユーザは、見た目の品質が高いウェブページを閲覧した場合に見た目に関するゲージのスコアが上昇し、内容に関するゲージのスコアが下降する。そのため、ユーザがどのようにしたらスコアが上



図 2 SimpleUI の図. SERP の右側に見た目と内容の品質に関するスコアを可視化したゲージを提示している。

下するかを直感的に理解できるインタフェースになっている。SimpleUI を用意した理由は、ReflectionUI と比較して、スコアの変動を通じその意味について内省を行う過程が、ユーザの情報精査にどのような影響を与えるかを検証するためである。

本稿では、提案インタフェースの効果を検証するため、クラウドソーシングを用いたオンラインユーザ実験を実施した。ユーザ実験の結果、以下のことが明らかになった。

- 提案インタフェースを用いたユーザが非大卒経験者の場合、内容の品質が高いウェブページを多く閲覧する傾向にあった。
- 提案インタフェースを用いたユーザが非大卒経験者の場合、ウェブページの内容の品質を重視する意識を有する傾向にあった。
- 提案インタフェースを用いたユーザが大卒経験者の場合、内容の品質が高いウェブページをあまり閲覧しない傾向にあった。

2 関連研究

2.1 ウェブ情報の品質とユーザ行動の関係

ウェブ情報の品質とユーザの行動の関係について研究はいくつか存在する。Suppanut [7] らは文書の信憑性と意見がウェブ検索ユーザに与える影響について分析した。結果、ユーザの持つ信念と相反するウェブページを提示したとき、より時間をかけることを明らかにした。Pogacar [8] らは、検索結果の情報の品質がユーザの検索行動に与える影響を調べるため、検索結果リストに含む情報の品質を操作し、ユーザの行動ログを分析した。結果、検索結果に誤った情報を多く含んでいたとき、ユーザは誤った判断をする可能性が高いことを明らかにした。Kammerer らは、ユーザが医療問題に関するウェブ上の情報を評価する際の行動を理解するため、提示する検索結果の品質を操作し、その行動を分析した [9]。結果、ウェブ上の情報は信憑性が高いと考えるユーザは、ウェブページの評価に時間をかけなくなることを明らかにした。

上記の研究は、ウェブページの品質をデザインや読みやすさといった「見た目の品質」、もしくはデータや証拠の有無といった「内容の品質」片方の側面のみに着目している。しかし、片方の側面だけでは、ユーザの閲覧したサイトの内容を正確に分析することが難しい。例えば、見た目の品質のみに着目した場合、一般のユーザが書いたデータや証拠の存在しないサイトで

あるか、専門家の書いたサイトであるかを判別することができない。本稿では、一般のユーザが学習する際、どのようなサイトを閲覧する傾向にあるか正確に知る必要があるため、ウェブページの品質を見た目と内容の品質の 2 点に着目している。

2.2 注意深いウェブ探索行動の促進

ウェブ検索ユーザに、注意深い検索行動を促すための情報アクセスシステムに関する研究はいくつか存在する。こうした研究は直接的にユーザに訴えかけるアプローチと間接的にユーザに訴えかけるアプローチが存在する。直接的なアプローチとして、Ennals らは、賛否両論のある文章にハイライト表示をし、賛成意見と反対意見の両面を提示するシステム DISPUTE FINDER を提案した [5]。また、Schwarz らは、ウェブサイトの信憑性を判断に関するスコアを可視化するシステムを作成した [10]。上記のシステムを用いることで、ユーザは検索結果の信憑性をより正確に判断できるようになった。

一方、間接的なアプローチは、直接ユーザにフィードバックを与えるのではなく、ユーザの批判的態度を喚起させるような方法を取っている。例えば、Yamamoto らは、検索システムのクエリ補完/推薦時に批判的思考を想起させるクエリを挿入し、批判的情報検索を促進するシステム QUERY PRIMING を提案した [11]。このシステムを用いることによって、ウェブ検索ユーザはクエリ修正を頻繁に行うようになり、データや証拠に基づいた情報を記したウェブサイトによく訪れることを明らかにした。また、Scott らは、他のウェブ検索ユーザの検索・閲覧行動の履歴を可視化するシステム SEARCH DASHBOARD を提案した [12]。ユーザ実験の結果、当該システムを用いたウェブ検索ユーザは自身の検索行動を改善することが可能であることを明らかにした。

本研究の提案するシステムは上記のように間接的なアプローチを採用している。具体的には、ユーザが閲覧したウェブページの品質を可視化することで、自身の閲覧傾向について内省を促し、注意深い情報探索を行ってもらおうという方法を採用した。

3 提案手法

3.1 提案インタフェース

提案インタフェースは、ユーザが閲覧したウェブページの品質をスコアによって可視化することで、ユーザの情報精査傾向の内省を促し、ページの見た目ではなく内容を重視した情報精査を行ってもらおうことを狙う。本稿では、上記のアイデアを実現するためのインタフェースとして、ReflectionUI と SimpleUI の 2 つを用意した。

ReflectionUI は図 3 に示すように、ユーザの情報精査傾向を表示するゲージが 1 つあり、スコアは内容の品質が高いページをどの程度閲覧したかを表している。なお、ユーザは SERP の下に見た目と内容の品質スコアが表示されているため、閲覧する前にどのようなウェブページであるかを知ることができる。スコアの決定方法については、節 4.4 にて後述する。

ReflectionUI は使用したユーザが内容の品質が高いページ

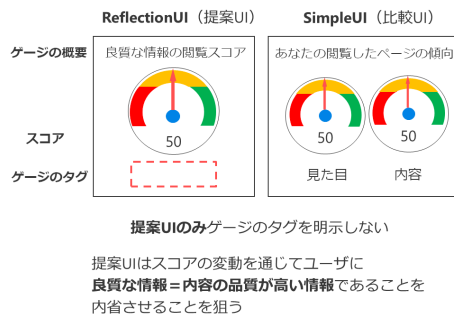


図 3 ReflectionUI と SimpleUI の概要

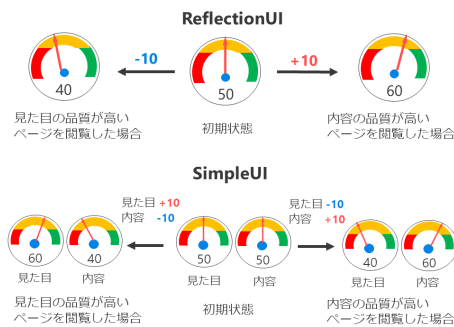


図 4 各 UI におけるスコア変動の動作

を閲覧した時のみ、ゲージのスコアが上昇する仕組みになっている。具体的には、図 4 に示すように、ユーザーが内容の品質が一定レベル以上のウェブページを閲覧するとスコアが 10 上昇し、そうではない場合はスコアが 10 下降する。スコアの変動幅は固定であり、最小値が 0 で最大値は 100 となっている。このように、スコアの上下変化を見せることで、ユーザーがウェブページの見た目ではなく内容の品質を重視したほうが良いことに気づかせることを狙う。重要な点は、ユーザーがゲージのスコアが上下する仕組みを陽に示していない点である。このような設計にした理由は、ユーザーがスコアの上下に囚われてしまい、その意味を考えなくなることを避けるためである。ゲーミフィケーションを利用したインタフェースは、ゲームをすることが目的になってしまい、その意味を考えないということが起こりうる。それゆえ、**ReflectionUI** は直感的にはわからないが、ユーザーが情報探索を行いながらスコアの上下を確認し、スコアを上昇させるための正しい行動を理解するという設計にした。

SimpleUI は図 3 に示すように、見た目の品質と内容の品質について、それぞれゲージを 2 つ用意した UI である。**SimpleUI** も **ReflectionUI** と同様に、ユーザーがウェブページを閲覧するごとにゲージのスコアが上下する。具体的には、図 4 に示すように、ユーザーが内容の品質が良いウェブページを閲覧すると内容の品質に関するゲージのスコアが 10 上昇し、見た目の品質に関するゲージのスコアが 10 下降する。**SimpleUI** を用意したのは、スコアの上下変化の理由が陽に示されていない **ReflectionUI** が、ユーザーの情報精査にどのような影響を与えるか検証するためである。

SimpleUI は見た目と内容のゲージがそれぞれあり、ユーザー

はどのようにしたらスコアが上下するか直感的に理解できるため、見た目ではなく内容を重視したほうが良いと内省を行う可能性が低い。それゆえ、**ReflectionUI** と **SimpleUI** を比較することで、情報探索時における内省の有無が、ユーザーの情報精査にどのように影響を与えるかを明らかにする。

3.2 仮説

本稿では、提案インタフェースの効果を検証するため、以下の仮説を立てた。

H1 提案インタフェースを使用するユーザーは、使用しないユーザーに比べて、情報を選び取るためにより多くの労力をかける。

H2 提案インタフェースを使用するユーザーは、使用しないユーザーに比べて、内容の品質が高いウェブページを多く閲覧する。

情報の吟味やトピックへの理解は批判的思考を要するため、提案 UI がすべてのユーザーに同様の効果が現れない可能性が高い。Yamamoto らによると、大卒経験者と非大卒経験者で批判的思考能力・態度に違いが生じることが明らかになっている [11] [13]。したがって、大卒経験者と非大卒経験者で提案インタフェースの効果が異なると考え、以下の仮説を立てた。

H3 **H1** および **H2** は、大卒経験者と非大卒経験者によってその程度が異なる。

4 実験方法

本章では、提案システムが、ユーザーの情報探索、情報精査に対する意識に与える影響について調査するオンラインユーザー実験について述べる。

4.1 検索タスク

実験協力者には検索トピックである「エイズ」と「糖尿病」に関して、ウェブ検索を通じて理解を深めてもらう検索タスクに取り組んでもらう。実験群は **Control** 群、**SimpleUI** 群、**ReflectionUI** 群の 3 群であり、実験協力者をいずれかの群にランダムに振り分けた。

4.2 実験手順

ユーザー実験は以下の手順で実施した。

- (1) ユーザー登録
- (2) 事前アンケート
- (3) 検索タスク
- (4) 事後アンケート

はじめに、実験協力者は [Lancers.jp](https://www.lancers.jp/)¹にてユーザー登録したあと、研究室で用意した実験用サイトに移動した。ユーザー登録を行ったあと、実験協力者をいずれかの群に振り分けた。その後、事前アンケートとして実験協力者には、検索トピックの「エイズ」もしくは「糖尿病」に関する理解度を測るためのクイズを行ってもらった。エイズに関するクイズは、中四国エイズセンター²が発行しているクイズを使用した。糖尿病に関するクイ

1 : <https://www.lancers.jp/>

2 : <https://www.aids-chushi.or.jp/>

ズは、日本糖尿病協会³が発行しているクイズを使用した。どちらもクイズの問題数は10問であり、実験協力者は問題に対して「はい」または「いいえ」で回答した。

次に、実験協力者には検索トピックの理解度を高めるための検索タスクに取り組んでもらった。すべての実験群は、筆者らが用意した検索結果リストを使用して検索タスクに取り組んでもらった。検索結果リストは節4.3にて後述する、ウェブページの品質を評価したデータセットのみが表示された。

実験協力者が検索タスクを終えたら、事前アンケートと同様のクイズを解いてもらい、検索タスクの前後で理解度がどのように変化したかを調べた。

最後に、実験協力者には情報精査の意識に関する質問、性別、年齢、学歴に答えてもらった。情報精査の意識に関する質問は、情報探索中にウェブページの内容もしくは見た目の品質をどの程度重視したかを5段階で答えてもらった。

4.3 ウェブページの見た目・内容の品質の事前調査

ウェブページの品質の評価については、Lancers.jp⁴を用いて、「エイズ」と「糖尿病」に関するウェブページ1000件をそれぞれ100名のクラウドワーカーに評価してもらった。なお、本稿では、ウェブページの品質について、デザインや読みやすさといった「見た目の品質」とデータや証拠の有無といった「内容の品質」の2つの観点に着目し、ワーカーに評価してもらった。ワーカーには1人あたり10件のウェブページの評価をしてもらい、謝金として40円を支払った。ワーカーに評価してもらうウェブページは、Microsoft Bing Web Search API⁵を用いて、検索トピックである「エイズ」と「糖尿病」の関連クエリによる検索を行い、日本語の検索結果のみを取得した。

ワーカーに提示した質問の詳細を以下に記す。検索結果が学習を行う際に有用であるかを確かめるための質問が上2つの項目、見た目の品質に関する質問は中2つの項目、内容の品質に関する質問は以下の2項目である。

- ウェブページの内容は「エイズ（糖尿病）」について学ばせて役に立つと思いませんか？(1: 役に立つと思う 2: 役に立たないと思う)
- ウェブページの内容は「エイズ（糖尿病）」に関連している内容でしたか？(1: 関連していると思う 2: 関連していないと思う)
- ウェブサイトのデザインはどの程度良いと感じましたか？(1: かなり悪い～5: かなり良い)
- ウェブサイトの文章はどの程度読みやすいと感じましたか？(1: かなり読みにくい～5: かなり読みやすい)
- ウェブサイトが公的機関（省庁、医療機関、学術組織）のデータや証拠がどの程度含まれていましたか？(1: 引用を含まない, 2: 公的機関の引用を1件は含む, 3: 公的機関の引用を3件以上含む)
- ウェブサイトの書き手はどの程度専門知識を有していま



図5 実験協力者に提示した検索結果リストの模式図

したか？(1: 書き手の情報が存在しない, 2: 書き手の情報は存在するが、専門家ではない, 3: 書き手の情報が存在し、専門家である。)

本稿では、ワーカーによる見た目もしくは内容の品質の評価が、ともに全体の平均値以上のウェブページを、見た目もしくは内容の品質が高いウェブページとする。

4.4 検索結果リスト

本実験の検索タスクでは、Google 検索⁵やYahoo 検索⁶といった一般的なウェブ検索エンジンが返す検索結果リストを模したものを提示した。検索結果リストには、4.3節で述べた品質を評価したウェブページをエイズと糖尿病でそれぞれ50件ずつ用意した。その中で、図5に示すように、見た目の品質が高いウェブページと内容の品質が高いウェブページを交互に表示した。順番はクラウドワーカーに評価してもらったスコアでソートし、上から順番に並べている。それに加えて、内容の品質が高いウェブページは見た目のスコアが1で内容のスコアが10、見た目の品質が高いウェブページは見た目のスコアが10で内容のスコアが1に固定した。ウェブページの品質のスコアをこのように設定したのは、ユーザに閲覧するウェブページの見たと内容の品質のどちらが良いかを直感的に判断してもらうためである。

なお、検索結果は一般的なウェブ検索の結果画面を模しているが、実験協力者がクエリの修正を行えないように設定した。各検索結果をクリックすると、実験開始前にあらかじめアーカイブしておいたリンク先のウェブページが表示されるようにした。さらに、各文書の閲覧時間を計測できるよう、アーカイブしたウェブページにJavaScriptのコードを埋め込んだ。文書中のハイパーリンクは無効化し、検索結果リストに表示された文書以外を閲覧できないように設定した。これによって、検索結果リストに掲載されたウェブページのみを対象に、確実にページ閲覧時間を計測することができる。

4.5 評価指標

仮説を立証するため、ユーザの検索行動、理解度の指標として以下のものを取得した。

- 検索結果一覧ページ閲覧時間（以下、SERP 閲覧時間）

3 : <https://www.nittokyo.or.jp/>

4 : <https://docs.microsoft.com/ja-jp/azure/cognitive-services/bing-web-search/>

5 : <https://www.google.co.jp/>

6 : <https://www.yahoo.co.jp/>

表 1 実験協力者の内訳

UI	学歴	
	非大卒経験者	大卒経験者
Control	18	30
ReflectionUI	24	28
SimpleUI	25	25

- ウェブページの閲覧時間
- 検索結果のクリックスルー

SERP 閲覧時間は、実験協力者が検索タスク中に SERP を閲覧した合計時間である。ウェブページの閲覧時間は、SERP からリンクされたウェブページの閲覧時間である。検索結果のクリックスルーは、実験協力者が SERP でクリックした検索結果の情報である。クリックスルー情報には、タイトル、概要文、URL、検索結果順位、ページの品質スコアが含まれる。

仮説 H1 を立証するため、提案システムを使用したユーザは使用しないユーザに比べ、SERP 閲覧時間が長く、内容の品質が高いページの閲覧時間が長くなるかどうかを検証する。また、仮説 H2 を立証するため、提案システムを使用したユーザは使用しないユーザに比べ、ウェブページの品質が高いページを多く閲覧するかどうかを検証する。

4.6 実験協力者

実験協力者はクラウドソーシングサイトの Lancers.jp を用いて、159 名を募集した。表 1 に記したとおり、150 名の実験協力者はランダムに Control 群、ReflectionUI 群、SimpleUI 群のいずれかの群に振り分けられた。実験終了後、謝金として 200 円を支払った。

5 結果

ユーザ実験の結果、合計 159 名の実験協力者から行動ログ、事後アンケートのデータを収集した。そのうち、検索タスクに真剣に取り組んでいないと思われる 9 名の実験協力者のデータを除外し、合計 150 名のデータを収集した。本章では、実験協力者のタスク中の行動データ、事後アンケートの分析結果について述べる。

本稿では、提案インタフェース（以下、UI 条件）と学歴を 2 要因とする分析を行った。UI 条件には Control 群、SimpleUI 群、ReflectionUI 群の 3 水準を設定し、学歴要因には大卒経験者と非大卒経験者の 2 水準を設定した。実験群の内訳は表 1 に記す。その上で、Control 群と比較して、SimpleUI 群と ReflectionUI 群にどのような違いがあるのかを分析した。

5.1 統計解析

ユーザ実験で収集した行動ログを解析するために、本研究では一般化線形混合モデル（GLMM）を用いた [14]。GLMM は介入による主効果と、実験協力者や検索トピックといった個体差による影響であるランダム効果を分離しながら分析を行うことができる。本稿では、GLMM をベイズ統計モデルに拡張した

モデルを用いて、行動データのモデリングを行った。ベイズ統計モデルは頻度主義的アプローチほど定着していないが、ターゲットとなるパラメータの事前分布を設定することで、データの不確実性を考慮したモデリングを行える。そのため、本稿ではベイズ統計モデルで拡張した GLMM を用いた。

本稿では、SERP 閲覧時間はワイブル分布、ページビュー数はポアソン分布、トピックに対する理解度と情報精査への意識は正規分布に従うと仮定した。

本稿では、提案インタフェースが及ぼす影響について調べるための指標として、95%最高密度区間（High Density Interval: HDI）を用いた。95%HDI はパラメータが確信度 95%で取りうる範囲を表したものであり、この区間にゼロを含まない場合、パラメータが有効であることを意味する（頻度主義統計における帰無仮説を棄却することに相当）。

5.2 SERP 閲覧時間

提案インタフェースを使用した実験協力者が、検索結果一覧リスト（SERP）に表示されたウェブページを選択する際、見た目と内容の品質の観点について、どの程度時間をかけて意識して選択したかを分析するため、SERP 閲覧時間を比較した。表 2 が示すように、Control 群と比較した場合の ReflectionUI と SimpleUI の結果は、UI 条件、学歴、交互作用の係数の 95%HDI にゼロが含まれていた（頻度主義統計における帰無仮説の棄却しないことに相当）。上記の結果は、SERP 閲覧時間に関して UI 条件による影響がないことを示唆している。

5.3 最大ページ閲覧時間

提案インタフェースを使用した実験協力者が、SERP に掲載されたウェブページをどの程度注視したかを分析するために、実験協力者がタスク中に閲覧した見た目もしくは内容の品質が高いウェブページの閲覧時間の最大値（最大ページ閲覧時間）を比較した。表 2 が示すように、Control 群と比較した場合の最大ページ閲覧時間に関する ReflectionUI と SimpleUI の結果は、UI 条件、学歴、交互作用の係数の 95%HDI にゼロが含まれていた。この結果は、最大ページ閲覧時間に関して UI 条件による影響がないことを示唆している。

5.4 ページビュー数

提案インタフェースを使用した実験協力者が、どのような品質のウェブページを閲覧したかを分析するため、タスク中に実験協力者が閲覧した見た目もしくは内容の品質が高いウェブページ数（ページビュー数）を比較した。表 2 が示すように、Control 群と比較した場合のページビュー数（見た目）に関する ReflectionUI と SimpleUI の結果は、UI 条件、学歴、交互作用の係数の 95%HDI にゼロが含まれていた。この結果は、見た目の品質が高いウェブページのページビュー数に関して UI 条件による影響がないことを示唆している。

一方、ページビュー数（内容）において、Control 群と比較した場合の ReflectionUI の結果は、交互作用の係数の 95%HDI にゼロを含んでいなかった（頻度主義統計における帰無仮説の棄却することに相当）。交互作用が確認されたため単純主効果

表 2 **Control** 群と比較したときの **ReflectionUI** と **SimpleUI** の GLMM の結果.

数値は 95%HDI の中央値と区間を表している. 95%HDI 区間にゼロを含まない数値は太字で示す.

検索指標	ReflectionUI			SimpleUI		
	UI	学歴	交互作用	UI	学歴	交互作用
SERP 閲覧時間	0.05	0.12	0.25	0.21	0.12	0.31
	[-0.33, 0.44]	[-0.55, 0.34]	[-0.36, 0.84]	[-0.20, 0.62]	[-0.55, 0.34]	[-0.31, 0.91]
最大ページ閲覧時間 (見た目)	-0.16	-0.16	0.36	-0.04	-0.16	0.13
	[-0.67, 0.36]	[-0.73, 0.42]	[-0.43, 1.14]	[-0.58, 0.49]	[-0.73, 0.42]	[-0.68, 0.91]
最大ページ閲覧時間 (内容)	-0.14	-0.40	0.46	0.08	-0.40	0.34
	[-0.61, 0.34]	[-0.93, 0.16]	[-0.29, 1.20]	[-0.43, 0.57]	[-0.93, 0.16]	[-0.46, 1.11]
ページビュー数 (見た目)	-0.23	-0.18	0.50	-0.22	-0.18	0.46
	[-0.67, 0.20]	[-0.66, 0.33]	[-0.16, 1.16]	[-0.65, 0.23]	[-0.66, 0.33]	[-0.19, 1.16]
ページビュー数 (内容)	-0.22	-0.33	0.73	-0.16	-0.33	0.56
	[-0.56, 0.14]	[-0.73, 0.07]	[0.17, 1.27]	[-0.55, 0.20]	[-0.73, 0.07]	[-0.01, 1.12]

事後アンケート	ReflectionUI			SimpleUI		
	UI	学歴	交互作用	UI	学歴	交互作用
見た目の重視度	0.01	$6.21e^{-3}$	$7.00e^{-3}$	-0.39	$6.21e^{-3}$	0.47
	[-0.62, 0.55]	[-0.60, 0.61]	[-0.87, 0.90]	[-0.88, 0.26]	[-0.60, 0.61]	[-0.45, 1.20]
内容の重視度	-0.08	-0.40	0.64	-0.17	-0.40	0.56
	[-0.49, 0.30]	[-0.85, 0.04]	[0.05, 1.28]	[-0.59, 0.22]	[-0.85, 0.04]	[-0.06, 1.19]

分析を行った結果, **ReflectionUI** を使用した非大卒ユーザと大卒ユーザに統計的有意差が確認された ($p < .05$). 図 6 に示すように, ユーザが非大卒経験者の場合, **ReflectionUI** 群は **Control** 群よりも, 内容の重視したウェブページをより多く閲覧する傾向にあった (6.71 vs. 3.61). 一方, ユーザが大卒経験者の場合, **ReflectionUI** 群は **Control** 群よりも, 内容の重視したウェブページを閲覧しない傾向にあった (4.04 vs. 5.28).

5.5 重視した観点

提案インタフェースがユーザの情報精査に対する意識にどのような影響を与えるか分析するため, 実験協力者が検索タスク中にウェブページの内容と見た目の品質をどの程度重視したかのアンケート回答を比較した. 表 2 が示すように, **Control** 群と比較した場合の見た目の重視度に関する **ReflectionUI** と **SimpleUI** の結果は, UI 条件, 学歴, 交互作用の係数の 95%HDI にゼロが含まれていた. この結果は, 見た目の重視度に関して UI 条件による影響がないことを示唆している.

一方, 表 2 が示すように, **Control** 群と比較した場合の内容の重視度に関する **ReflectionUI** との結果は, 交互作用の係数の 95%HDI にゼロが含まれていなかった. 交互作用が確認されたため単純主効果分析を行った結果, **ReflectionUI** を使用した非大卒ユーザに統計的有意差が確認された ($p < .05$). 図 7 に示すように, ユーザが非大卒経験者の場合, **ReflectionUI** 群は **Control** 群よりも, ウェブページを選ぶ際に内容の品質を重視する傾向にあった (4.17 vs. 3.60).

6 考察

6.1 仮説検証

本稿では, ユーザのウェブ検索にかけた労力に関する仮説

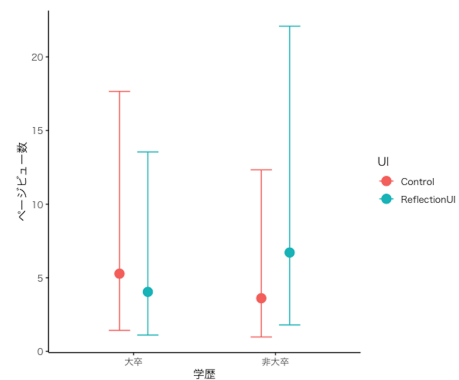


図 6 ページビュー数に対する UI 条件と学歴の効果の推定値. 赤色の線は **Control** 群, 青色の線は **ReflectionUI** 群を表し, バーは信用区間を示す.

H1 を検証するため, SERP 閲覧時間, 最大ページ閲覧時間について分析した. 分析の結果, これらの指標には UI 条件が影響することは認められなかった. 想案では **ReflectionUI** を使用したユーザは, 品質の良い情報は何かと注意深く考えて情報探索を行おうとする結果, SERP 閲覧時間が長くなると予想したが, そのような結果は得られなかった. このような結果になった理由は, SERP に表示したウェブページの品質を表すスコアにあったと考えられる. 品質を表すスコアは, ユーザがウェブページの見た目と内容のどちらの品質が高いか直感的にわかるように表示した. それゆえ, 検索結果の概要分やページ文書本体を注意深く読まなくても, ユーザはどのページが品質の高いウェブページであるかを考える必要がなかったとも考えられる. したがって, SERP 閲覧時間には UI 条件による影響が認められなかったと考えられる.

想案では **ReflectionUI** を使用したユーザは, 内容の品質が

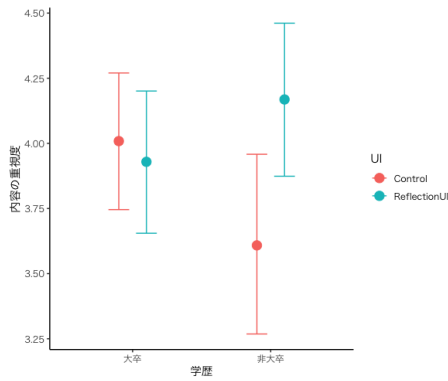


図7 内容の重視度に対するUI条件と学歴の効果の推定値。赤色の線はControl群、青色の線はReflectionUI群を表し、バーは信用区間を示す。

高いウェブページを注視しようとする意識を持つため、内容の品質が高いウェブページの閲覧時間が長くなると予想した。しかし、そのような結果は得られなかった。これはSERP閲覧時間と同様に、ウェブページに表示したスコアに原因があると考えられる。ReflectionUIを使用したユーザは、内容の品質が高いウェブページを見ようという意識があったとしても、スコアの表層的な情報によってその品質を判断し、内容を注視する意識が薄れてしまった可能性がある。それゆえ、最大ページ閲覧時間にはUI条件による影響が認められなかったと考えられる。以上のことから、仮説H1は支持されなかったといえる。

本稿では、ユーザが閲覧するウェブページの品質に関する仮説H2を検証するため、ページビュー数について分析した。分析の結果、ユーザが非大卒経験者の場合、ReflectionUI群はControl群よりも、内容の品質が高いページをより多く閲覧する傾向にあった。さらに、ユーザが情報探索中にウェブページの内容の品質をどの程度重視したか分析した。分析の結果、ユーザが非大卒経験者の場合、ReflectionUI群はControl群よりも、ウェブページの内容の品質を重視する意識を有する傾向にあった。上記の結果から、非大卒経験者のユーザがReflectionUIを使用した場合、Control群よりも、ウェブページの内容の品質を重視していたといえる。大卒経験者のユーザは批判的思考を有している可能性が高いため、ReflectionUIを使用した場合、Control群よりも、内容の品質が高いウェブページを多く閲覧すると筆者は予想した。しかし、実際には予想と反対の結果であった。

このような結果になった理由は、ユーザの批判的思考能力の違いにあると考えられる。Kusumiらによると、情報を適切に判断するためには、言語能力や推論力といった批判的思考スキルだけでなく、批判的であろうとする批判的思考態度が必要とされている[6]。大卒経験者はもともと批判的スキル・態度を有している可能性が高いため、ReflectionUIのようなツールに頼らずとも、自らの能力によって批判的に情報探索を行い、ほぼほどに検索タスクを終えると考えられる。一方、非大卒経験者のユーザは、大卒経験者のユーザより、ものごとを批判的に考える経験が少ないため、ウェブ情報を批判的に探索しない可

能性が高い。しかし、ReflectionUIのようなツールによって、自身の批判的思考態度が喚起され、批判的に情報探索を行おうとする意識が強まったと考えられる。それゆえ、非大卒経験者のユーザはReflectionUIを使用した場合、Control群よりも批判的に情報探索を行う傾向にあったと考えられる。以上のことから、仮説H2は支持されたといえる。また、学歴に応じて提案インタフェースの効果が異なったため、仮説H3は部分的に支持されたといえる。

6.2 情報精査傾向の内省を促すインタフェースの設計

今回のユーザ実験の結果を踏まえ、ユーザの情報精査傾向の内省を促すためのインタフェースの設計について考察する。このようなインタフェースを作成するには2つの点が必要であると考えられる。1つ目はインタフェースの意図がユーザが直感的にわからないという点である。今回の実験結果より、ReflectionUI群は非大卒経験者ユーザの場合、Control群よりもウェブページの内容の品質を重視している傾向にあった。しかし、SimpleUI群にはそのような傾向は確認されなかった。ReflectionUIとSimpleUIの違いは、インタフェースの意図がユーザが直感的にわからないという点にある。したがって、ユーザが初めてインタフェースに接した際、その意図するところを直感的にわからないように設計することが、ユーザの情報精査の内省を促すために重要であるといえる。

2つ目は、ユーザの批判的思考に応じてインタフェースの設計を変えるという点である。今回の結果から、ReflectionUIを使用した非大卒経験者は、ウェブページの内容の品質を重視する傾向にあることが明らかになった。しかし、大卒経験者がReflectionUIを使用した場合、内容の品質が高いページを閲覧しない傾向にあった。このことから、ものごとを批判的に考える経験の差が、提案インタフェースの效果に違いを及ぼしたと考えられる。そのため、ものごとを批判的に考える経験の少ないユーザにとって、今回のようなゲーミフィケーションを利用したインタフェースは、批判的なウェブ探索を促すために有用であるといえる。しかし、もともと批判的思考を有しており、普段からそのような意識を持ってウェブ探索を行っているようなユーザには、提案インタフェースはあまり有用ではないといえる。それゆえ、ユーザの内省を促すようなシステムを設計する場合、ユーザの批判的思考能力に応じてインタフェースの設計を変える必要があると考えられる。

6.3 改善点

今回の結果を踏まえて改善すべき点は2つある。1つ目はウェブページに提示したウェブページの品質のスコアである。今回はユーザにどのウェブページが見た目もしくは内容の品質がどうか直感的に判断してもらうため、内容と見た目の品質のスコアは一律で固定にした。しかし、そのようにしたことによって、ユーザはウェブページの見たと内容のどちらの観点を重視しようか考えることに時間をかけなくなってしまった可能性がある。したがって、ウェブページの品質の自動推定技術[15]などを用いて、手動でタグ付をしないようにする必要があると考え

られる。

2つ目は実験設計である。今回はクラウドソーシングを用いたオンラインユーザ実験を行い、行動ログとアンケートによる分析を行った。しかし、実際にユーザが **ReflectionUI** を使用してどのように考えたかを正確に分析するには不十分であったと考えられる。それゆえ、自由記述のアンケートを行ったり、実際にユーザにインタビューを行うなどして正確に分析する必要がある。また、今回のユーザ実験では短期間であったが、提案インタフェースの効果の持続性について調べるため、長期に渡ってユーザ実験を行う必要がある。さらに、提案インタフェースがユーザの批判的思考を持続的に高めたかどうかを分析する必要がある。具体的には、実験協力者に提案インタフェースを使用したのち、提案インタフェースを使用せずに同様のタスクを行ってもらい、ユーザが内容の品質を重視するようになったかを確かめるためのユーザ実験を行う必要がある。

7 おわりに

本稿では、ウェブ検索ユーザの情報精査を促すため、ユーザが閲覧したページの品質を見た目と内容の観点から可視化し、ユーザのウェブ検索行動を内省させるためのインタフェースを提案した。提案インタフェースの効果を検証するため、クラウドソーシングを用いてオンラインユーザ実験を実施した。ユーザ実験の結果、提案インタフェースを使用したユーザが非大卒経験者の場合、使用しないユーザよりも、情報探索の際にウェブページの内容の品質を重視し、内容の品質ウェブページを多く閲覧する傾向にあった。一方、提案インタフェースを使用したユーザが大卒経験者の場合、使用しないユーザよりも、内容の品質ウェブページを閲覧しない傾向にあった。これらの結果は、提案インタフェースの効果が、ユーザの批判的思考能力に依拠して異なることを示唆している。また、非大卒経験者のように批判的にものごとを考える経験の少ないユーザにおいて、提案インタフェースのように自身のウェブ検索行動を内省させることは、情報精査を促すために有効であったことを示唆している。

今後は長期間にわたってユーザ実験を行い、提案インタフェースの使用感やユーザのウェブ検索行動や意識に与える影響について詳細に分析する必要がある。

謝 辞

本研究は JSPS 科研費 JP18H03244, 21H03554, 21H03775 の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] Elizabeth Sillence, Pam Briggs, Lesley Fishwick, and Peter Harris. Trust and mistrust of online health sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, p. 663–670. ACM, 2004.
- [2] Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Trans. Comput.-Hum. Interact.*, Vol. 18,

No. 1, May 2011.

- [3] Charles L. A. Clarke, Eugene Agichtein, Susan Dumais, and Ryan W. White. The influence of caption features on click-through patterns in web search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, p. 135–142, 2007.
- [4] Chee Wee Leong and Silviu Cucerzan. Supporting factual statements with evidence from the web. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, p. 1153–1162, New York, NY, USA, 2012. Association for Computing Machinery.
- [5] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. Highlighting disputed claims on the web. In *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)*, p. 341–350, 2010.
- [6] Takashi Kusumi, Rumi Hirayama, and Yoshihisa Kashima. Risk Perception and Risk Talk: The Case of the Fukushima Daiichi Nuclear Radiation Risk. *Risk Analysis*, Vol. 37, No. 12, pp. 2305–2320, 2017.
- [7] Suppanut Pothirattanachaiikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. Analyzing the effects of document's opinion and credibility on search behaviors and belief dynamics. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM 2019)*, p. 1653–1662, 2019.
- [8] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR '17*, p. 209–216, New York, NY, USA, 2017. Association for Computing Machinery.
- [9] Yvonne Kammerer, Ivar Bråten, Peter Gerjets, and Helge I. Strømsø. The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, Vol. 29, No. 3, pp. 1193–1203, 2013.
- [10] Julia Schwarz and Meredith Morris. Augmenting web pages and search results to support credibility assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, p. 1245–1254, New York, NY, USA, 2011. Association for Computing Machinery.
- [11] Yusuke Yamamoto and Takehiro Yamamoto. Query priming for promoting critical thinking in web search. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR 2018)*, p. 12–21, 2018.
- [12] Scott Bateman, Jaime Teevan, and Ryan W. White. The search dashboard: How reflection and comparison impact search behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2012)*, p. 1785–1794, 2012.
- [13] Yusuke Yamamoto, Takehiro Yamamoto, Hiroaki Ohshima, and Hiroshi Kawakami. Web access literacy scale to evaluate how critically users can browse and search for web information. In *Proceedings of the 10th ACM Conference on Web Science (WebSci 2018)*, p. 97–106, 2018.
- [14] D. Barr, R. Levy, C. Scheepers, and H. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, Vol. 68, No. 3, pp. 255–278, 2013.
- [15] Dara Bahri, Yi Tay, Che Zheng, Don Metzler, Cliff Brunk, and Andrew Tomkins. Generative models are unsupervised predictors of page quality: A colossal-scale study. 2021.