

批判的なウェブ検索を促進するクエリプライミング

山本 祐輔^{1,a)} 山本 岳洋^{2,b)}

受付日 2018年6月9日, 採録日 2018年9月27日

概要: 本稿では, ウェブ検索中のユーザに注意深い情報探索を暗黙的に促す「クエリプライミング」を提案する. クエリプライミングは, 批判的思考を喚起し注意深い情報探索や意思決定を促進するようなキーワードを, クエリ補完やクエリ推薦時に提示する. クエリプライミングの有効性を検証するために, 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った. 被験者の情報探索ログ分析および実験のアンケート調査の結果, 以下の傾向が明らかになった: (1) クエリプライミングが実装されたウェブ検索エンジンを用いた被験者はセッション中の検索回数が増え, 検索結果一覧ページを何度も見直すようになる. (2) クエリプライミングによって, 証拠を重視してウェブページを検索・閲覧する行動が促進される. (3) クエリプライミングの効果は被験者の学歴に依存する. 本研究で得られた知見は, 注意深い情報探索を活性化させる検索インタラクションの設計に寄与することが期待される.

キーワード: ウェブ検索, 批判的思考, プライミング効果, ヒューマンファクター

Query Priming for Promoting Critical Web Search

YUSUKE YAMAMOTO^{1,a)} TAKEHIRO YAMAMOTO^{2,b)}

Received: June 9, 2018, Accepted: September 27, 2018

Abstract: In this paper, we propose a novel method, *query priming*, to activate careful user information seeking during web search process. Query priming employs query auto-completion (QAC) and query suggestion (QS) to show search terms that stimulate critical thinking and encourages careful information seeking and decision making on the web. We conducted a user study using a Japanese crowdsourcing service. Through the user study, we found the followings: (1) Participants using a search user interface with query priming, issued more queries and (re-)visited search engine result pages more frequently. (2) Query priming promoted web page selection targeted at evidence-based decision making. (3) The query priming effect varied relative to participant educational background.

Keywords: web search, critical thinking, priming effect, human factor

1. はじめに

ウェブが重要な知識基盤の1つになっている一方で, ウェブにアップロードされた情報の信憑性が社会問題になっている. ウェブ情報にはしばしば誤情報が混在している. 近年では, フェイクニュースのように, ソーシャルメディア上に誤情報が意図的に流される事態も生じている. このよ

うな状況にもかかわらず, ウェブ情報の信憑性について疑問を抱いたことがないというユーザが相当数存在することが報告されている [11], [18], [19]. ユーザが誤った情報を鵜呑みにして被害を被らないようにするためにも, 信憑性の高いウェブ情報の取得を支援する情報アクセスシステムの研究開発が望まれる.

これまで, 信憑性の高いウェブ情報の取得を支援するアプローチとして, 証拠情報の検索システム [14] や反証が存在する情報の検知システム [7], [25], 信憑性判断時に重要となる指標に沿ったスコアの可視化システム [26] など, 様々なアイデアが提案されてきた. これらのシステムは, 適材適所に使うことで非常に有用である. しかし, ユー

¹ 静岡大学
Shizuoka University, Hamamatsu, Shizuoka 432–8011, Japan

² 京都大学
Kyoto University, Kyoto 606–8501, Japan

a) yusuke.yamamoto@acm.org

b) tyamamot@dl.kuis.kyoto-u.ac.jp

ザが信憑性に注意を払わず情報を精査しようとする意志がなければ、前述したシステムに必要性を感じない可能性がある。また、人は自分の意見や信念と対立する情報が存在することを知っていても、事前に抱いている信念や意見、先入観を支持する情報を優先的に選ぶ傾向がある [15]。このようなケースでは、前述の支援システムが適切に使われない可能性もある。

本研究では、検索ユーザにある性質を持つ語を提示することで注意深い情報精探を促進する、クエリプライミングという手法を提案する。提案するクエリプライミングは、認知科学で明らかにされているプライミング効果現象に基づく手法である。プライミング効果とは、先行する刺激（以下、プライムと呼ぶ）によって、その後に発生する認知処理が促進または抑制される現象である [12]。プライミング効果を示す有名な例として、Bargh らのフロリダ実験があげられる [3]。この実験では、被験者である学生は2グループに分けられ、片方のグループには「忘れっぽい」、「禿げ」、「ごま塩」、「シワ」など高齢者を連想させる語を使って短い文章を書くよう指示が与えられた。その後、各グループの歩行速度を計測したところ、高齢者を連想させる語を用いて作文を行ったグループはそのような語を使わずに作文したグループに比べ、歩行速度が有意に速かったということが報告されている。

情報の質や信憑性を評価するには、批判的思考が必要とされている [2], [28]。本稿では、批判的思考や注意深い情報探索を促進するような語であるプライム語を設計する。さらに、検索エンジンのクエリ補完時、クエリ推薦時にプライム語を提示しクエリプライミングを行う検索インタフェースを提案する。図 1 は、クエリプライミングを実装したクエリ補完と従来のクエリ補完の比較を示したものである。図が示しているように、提示するプライム語は検索タスクの補完情報として提示しても違和感のないものを選択し、検索行動を妨げることなくクエリプライミングの効果を発現させられるよう、検索インタフェースの中に溶け

込ませる。本稿で提案するクエリプライミングは、任意の検索セッションにおいて検索行動を干渉しない形で批判的思考態度を喚起し、注意深い情報探索を促進することが期待される。

以下、本稿の構成を記す。2章では関連研究について整理する。3章では、クエリプライミングの要件と設計方針について整理する。4章では、クエリプライミングで用いるプライム語の収集方法について述べる。5章では、設計したクエリプライミングを用いたユーザ実験について説明し、6章ではクエリプライミングの効果について分析を行う。7章では、学歴とクエリプライミングの効果の関係など、実験を通して明らかになった知見や課題について整理し、結びとする。

なお、本稿は、the 3rd ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2018) にて著者らが発表した論文 [27] とほぼ同様の内容である。

2. 関連研究

2.1 ウェブ情報の信憑性判断支援

信憑性判断支援のアプローチとして、ツイート情報、言説、ウェブページなど、様々なウェブ情報の正確性や信憑性を数値化するアルゴリズムが提案されつつある。Pasternack らは、入力された言説に類似する言説、矛盾する言説を集約することで言説の信憑性を評価する Latent Credibility Analysis (LCA) アルゴリズムを提案している [21]。Dong らは、誤った事実や主張を含まないウェブページほど正確であると仮定し、ウェブページの正確さを評価するアルゴリズムを提案している [6]。Castillo らは、ニュース記事に対する tweet (つぶやき) や記事の共有行動を解析することで、Twitter 上で伝播するニュース記事の信憑性を評価するアルゴリズムを提案している [5]。

非数値情報を提示することによって信憑性判断を支援することに焦点を当てたシステムもいくつか提案されている。Leong らはユーザがウェブページを閲覧中に疑わしい文の信憑性を検証できるよう、証拠情報の検索アルゴリズムを提案している [14]。Ennals らは閲覧中のウェブページの内容に反証が存在する場合、その箇所をハイライトするシステム DISPUTE FINDER を提案している [7]。

2.2 ウェブ探索時の態度と認知バイアス

ウェブ上には検証されていない情報が多数存在するにもかかわらず、相当数のウェブユーザが不確かなウェブ情報を鵜呑みにしてしまっているということが報告されている。たとえば、Nakamura らの調査では、ウェブ検索ユーザの半数以上はウェブ検索エンジンが提示する検索結果の上位に含まれるページであれば信用できると考えているということが報告されている [19]。Morris らは、ソーシャルネットワーク上では誤った情報が蔓延しているにもかか

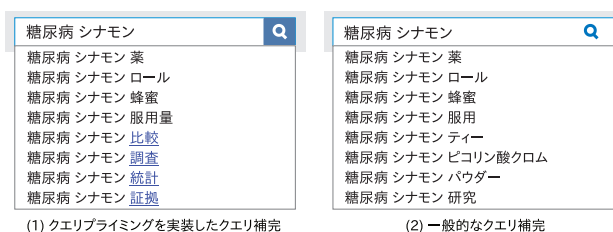


図 1 クエリプライミングが実装されたクエリ補完と従来のクエリ補完の比較 (違いを分かりやすくするために、論文上ではプライム語に青色の下線を引いている)

Fig. 1 Comparison of query auto-completion (QAC) with query priming and conventional QAC (prime terms are underlined in this paper so that readers can identify them as manipulated).

ならず、検索エンジンが返す情報よりもソーシャルネットワーク上の情報を信用しているユーザが多いことを明らかにしている [18].

疑わしい情報の存在を意識していたとしても、人は誤ったヒューリスティックの使用、すなわち認知バイアスの影響で信憑性に関する誤判断を下してしまうことがある [12]. Jeong らは、ウェブユーザが特定のドメイン上で公開されているウェブページを信用してしまう傾向、ドメインバイアスの存在を明らかにしている [10]. White らは検索トピックに対する事前信念と検索行動の関係について分析を行っており、ユーザが検索トピックに関して強い事前信念を持っている場合は、ウェブ検索・閲覧をして様々な情報を見たとしても、事前信念が修正されることは少ないことを明らかにしている [23].

2.3 良質な検索行動の促進

本稿で提案するクエリプライミングと同様に、ユーザの検索行動の質を高める方法に関する研究事例もいくつか存在する. Harvey らは、適合文書を導くクエリ例を提示することで、ユーザのクエリ生成スキルを高められることを明らかにしている [9]. Bateman らが開発した SEARCH DASHBOARD は、検索行動履歴を集約提示することで検索行動に対する内省を促すシステムである. Bateman らの実験によると、SEARCH DASHBOARD を使用したユーザは、自身の検索パフォーマンスを改善すべく検索行動を改めるようになったということが報告されている [4]. これらの研究は、ユーザの検索スキルを高めるために、明示的なフィードバックを返すアプローチを採用している.

一方、検索体験の質を向上させるために、暗黙的にユーザの検索行動を変容させることに焦点を当てた研究も存在する. Yamamoto らは閲覧中の検索結果やウェブページ中の反証が存在する語をハイライトすることで、ユーザは検索結果の選択により長い時間をかけるようになることを明らかにしている [25]. Agapie らはユーザが長いクエリを入力すると検索ボックスが光る検索インタフェースを提案している. 実験によると、提案インタフェースを用いたユーザは適合性の高いウェブページが得られるよう、長いクエリを入力するようになったことが報告されている [1]. 本稿で提案するクエリプライミングは注意深い情報探索に誘導するという意味で、暗黙的な行動変容アプローチの一種と考えることができる.

3. クエリプライミングの設計

3.1 ウェブ検索と批判的思考

クエリプライミングの目的は、ウェブ検索中のユーザに批判的思考に基づく情報探索を促すことにある. Ennis は、批判的思考を何を信じ何を行うべきかを決定するために行う、論理的で省察的な思考と定義している [8]. また

Ennis は批判的思考を実行できる人物は以下の行動をとろうとする意志を有するとしている：根拠の探索、俯瞰的な考察、複数の可能性の考慮、演繹的推論. 本稿では、検索ユーザが批判的思考を行える人物であれば、正確で信用できる情報をウェブから収集するために、情報リテラシーの研究者や図書館司書が重要視する以下の行動をとると仮定する [16].

- (1) 検索に十分時間をかける.
- (2) 意志決定に必要な情報を得るために複数回検索を行う.
- (3) 比較のために複数のウェブページを閲覧する.
- (4) コンテンツ作成者の専門性や、参照情報の有無、情報の鮮度など、ページに書かれている内容の質を担保する証拠情報を確認する.
- (5) 確実な意志決定を行うための証拠情報を集める.

我々はウェブ検索プロセスにおいて、上記の振舞いを促進するようなクエリプライミングを行う検索インタフェースを提案する.

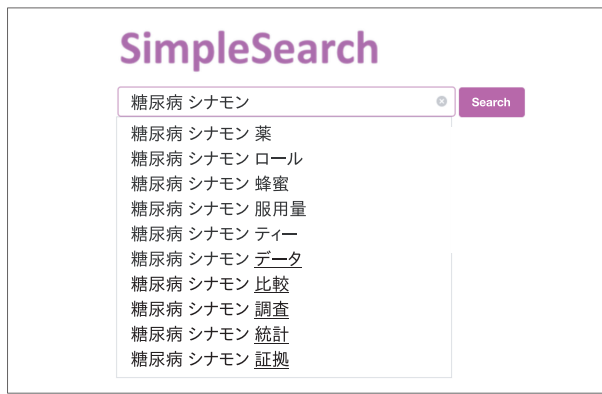
3.2 ウェブ検索におけるクエリプライミングの設計

本稿で提案するクエリプライミングは、プライミング効果の1種であるイデオモータ効果 [3] に着目したものである. イデオモータ効果とは、プライミング効果の1種で先行刺激によってある心理概念が喚起され、それによってその後発生する行動が変化するという現象である. 基本的なアイデアは、ウェブ検索ユーザがクエリを入力中あるいは入力直後のタイミングで、批判的思考態度を喚起し注意深い情報探索を促す語（以下、プライム語と呼ぶ）をユーザに提示するというものである. 検索エンジンの利便性を損なわずに注意深い情報探索を促すには、提示するプライム語は少なくとも以下の要件を満たす必要がある.

第1に、プライミング効果によって批判的思考態度を喚起し行動変容を引き起こすには、プライム語から批判的思考の概念を連想できなければならない. 第2に、プライム語は検索タスクの補完情報として提示しても違和感がないようにする必要がある. 提示するプライム語が批判的思考を喚起するものであったとしても、検索タスクと無関係な語が提示されるようであれば、検索行動の妨げになる. 第3に、プライム語が必ず検索ユーザの目に触れるよう、検索システムにプライム語の提示機会をうまく設ける必要がある. 可能なら、複数回目に触れる機会があることが望ましい.

上記の要件を満たすべく、以下の方針でクエリプライミング機能の設計を行う. 批判的思考態度といっても具体的なイメージは人によって様々であると考えられる. 本研究では要件1を満たすために、平山らの研究を参考にプライム語を設計する [29]. 平山らの研究によると、批判的思考態度として「論理的思考の自覚*1」、「探究心*2」、「客観

*1 論理的思考の重要性を認識し、論理的思考を自覚的に活用しようとする態度.



(1) ページ画面におけるクエリプライミング(クエリ補完)



(2) SERPにおけるクエリプライミング(クエリ推薦)

図 2 (1) クエリ補完におけるクエリプライミングの例, (2) クエリ推薦におけるクエリプライミング例 (プライム語とそうでない語の違いを分かりやすくするために、論文上ではプライム語に下線を引いている)

Fig. 2 Screenshot of (1) query auto-completion with query priming and (2) query suggestion with query priming (prime terms are underlined in this paper so that readers can identify them as manipulated).

性*3],「証拠の重視*4」の4つの態度が見い出されている。本研究では、これら4つの態度を連想する語を批判的思考を連想する語として収集する。要件2については、批判的思考態度を持った人物が汎用的に利用すると思われるクエリ語を収集することで対応する。また、批判的思考態度を喚起する語のうち、クエリの一部として用いられても違和感のない語を選択することで対応する。要件3については、プライム語を提示する機会としてクエリ補完・推薦機能に着目する。クエリ補完機能 [17] やクエリ推薦機能 [13] は、検索ボックス付近で検索支援情報を提供する機能として、ウェブ検索エンジンユーザに馴染みがあるものである。ウェブ情報検索において、ユーザが必ず行う行為はクエリの入力・修正である。それゆえ、クエリ入力中あるいは入力直後にプライム語の提示を行うことで、プライム語を目視する機会を確実に設けることができる。

図 2 は、クエリプライミングを実装したクエリ補完、クエリ推薦結果のスクリーンショットである。図が示しているとおり、プライム語はクエリ補完・推薦リストの中で提示される。このような方法でクエリプライミングを行うことで、注意深いウェブ情報探索が促進されることが期待される。

3.3 リサーチクエスチョン

クエリプライミングによって批判的思考態度が喚起された場合、ユーザの検索・閲覧行動は慎重で注意深い情報探索に向けて、表層的にも本質的にも変化することが期待される。また、クエリプライミング機能を実装した検索インタフェースの使用をやめた後でもクエリプライミングの効

果が持続するならば、そのような検索インタフェースは批判的思考の訓練にも活用できる可能性がある。一方、クエリプライミングの効果はすべてのユーザに発現しない可能性もある。ある研究グループによると、研究プロジェクトやグループワークへの参加、小論文形式のレポート課題への取り組みといった大学での学習活動が批判的思考能力を高めると報告されている [20], [22]。この知見から、クエリプライミングの効果はユーザの批判的思考能力に依存するとも考えられる。

本研究ではクエリプライミングの効果を検証するために、以下のリサーチクエスチョンを設定する。

- **RQ1**: クエリプライミングは検索タスクにかかる時間、クエリの発行回数、ページ訪問数など、検索・閲覧行動に影響を及ぼすか。
- **RQ2**: クエリプライミングは意思決定の参考とするウェブページを選択する際の観点に影響を及ぼすか。
- **RQ3**: クエリプライミングが検索ユーザに何らかの影響を与える場合、実装された検索システムの使用をやめた後も、その影響は持続するか。
- **RQ4**: 大卒経験が批判的思考能力に関連していると仮定した際、大卒経験とクエリプライミングの効果に相関関係はあるか。

4. クエリプライミングで用いる語の収集

本研究では、クエリ推薦・補完時に提示するプライム語をクラウドソーシングを用いて収集、評価した。以下、手順の詳細と得られたプライム語について述べる。

4.1 プライム語候補の収集

検索エンジンユーザの批判的思考態度を喚起させるためにクエリ補完・推薦時に提示するプライム語は、当該態度

*2 開かれた心で様々な情報を求めようとする態度。
 *3 主観にとらわれず客観的に考えようとする態度。
 *4 適切な証拠を求め、それに基づき判断しようとする態度。

表 1 プライム語収集タスクで用いた検索トピック

Table 1 Search topics to collect prime term candidates.

検索トピック
引越越し業者 契約, インターネット回線 契約, 癌 治療法 ダイエット 方法, 海外留学 検討, オンライン英会話 検討 電子レンジ 購入, テレビ 購入, 株式投資 資産運用 クレジットカード 申し込み

を持った人物やその行動を連想させる語であると同時に、入力クエリと組み合わせても違和感を感じない語である必要がある。そこで、下記に述べる 2 種類の方法でプライム語候補を収集した。

まず、批判的思考態度を持った人物から連想されるキーワードをクラウドソーシングサービス Lancers.jp^{*5}を用いて収集した。本収集タスク（収集タスク 1）では、3.2 節で述べた批判的思考態度を構成する 4 種類の態度のいずれかを取り上げ、そのような態度を持った人物像から連想される人物の性格、あるいはその人物がとりうると思われる行動に関するキーワードをクラウドワーカに 3 つあげさせた。この方法によって、74 名のワーカから 401 個のキーワードを収集した。

別の方法として、批判的思考態度を持った人物が任意のトピックについてウェブ検索を行う際、クエリ入力時に利用すると思われるキーワードについて、Lancers.jp を用いて収集した。本収集タスク（収集タスク 2）では、4 種類の批判的思考態度のいずれかを有した人物が表 1 に載せたトピックに関してウェブ検索する場面について、クラウドワーカに想像させた。そのうえで、各ワーカには、批判的思考態度を持った人物が検索トピックとともに AND オペレータとともにクエリとして入力しうる語を 3 つあげてもらった。この方法によって、54 名のワーカから 317 個のキーワードを収集した。

収集したキーワードの一部は巻末の付録に掲載した。

4.2 プライム語候補の評価

収集タスク 1 および 2 で収集したプライム語候補群から批判的情報探索を促すクエリプライミングに用いるプライム語を選定するため、再度クラウドソーシングを用いた評価を行った。

本評価タスクでは、収集したプライム語候補を構成要素とするクエリで検索を行うユーザをクラウドワーカに想像させた。その後、想像した検索ユーザが批判的思考ができる人物であると感じる程度を評価させた。最終的に、ワーカの評価値の平均値をプライム語候補から批判的思考態度を連想する度合いと見なした。

本評価タスクでは、4 種類ある批判的思考態度のそれぞれについて、収集タスクで出現頻度が高かった上位 20 件

の語を抽出し、それらを評価対象とした。タスクは以下の手順で行われた。

- (1) 4 種類の批判的思考態度から評価対象とする態度 d をランダムに選び、ワーカに割り当てる。
- (2) d に関して 4.1 節で収集されたプライム語候補 20 件について、以下の手順を行った（プライム語を p とする）。
 - (a) クラウドワーカに、 p を含むクエリを入力した検索ユーザを想像させる。
 - (b) 想像した検索ユーザの批判的思考態度 d の度合いを 5 段階のリッカート尺度で評価させる。

以下は、タスクで用いた説明文の一例である。

ある人物がトピック $\langle topic \rangle$ についてウェブ検索しているときに、「 $\langle topic \rangle$ & 比較」という検索ワードを使ったとします。この検索ワードから、検索中の人物が「客観的な視点を持った人物」であると感じますか？ 5 段階で評価してください（2：かなり感じる～～2：まったく感じない）。なお、この人物はどんなトピックを調べる時も検索ワードに「比較」という語を付け足すことが多い人物と考えてください。

本評価タスクでは、4 種類ある批判的思考態度ごとに 50 名のワーカを割り当てた。表 2 に、本評価タスクで高評価を集めたプライム候補語の上位 5 件を記す。評価が高い語と低い語を比較した結果、評価の低い語は批判的思考を持った人物が持つ「性格」に関する語が多い傾向にあった（例：頭脳明晰、落ち着き、柔軟、冷静）。評価が低い語の中には、「旅行」、「価格」のように特定のトピックに関連する語も少なからず含まれていたが、トピックに関連する語の多くは「収集タスクにおける出現頻度上位 20 件」という制約によってある程度除外されていることを確認した。結果として、ワーカの評価値の平均値を用いることで、トピック非依存で、検索ワードとして追加しても違和感のない語が上位に集まった。最終的に、各批判的思考態度のスコア上位 3 件のプライム語候補から類似するものを除いた 10 個^{*6}について、クエリプライミングに用いるプライム語とした（表 2 中の下線付きの語）。

5. 実験

本章では、クエリプライミングの効果を分析するために行ったユーザ実験について述べる。

5.1 被験者

クラウドソーシングサービス Lancers.jp を用いて、計 200 名の被験者を募集した。200 名のうち 82 名はタスクが未完了であったり、実験中にこちらが用意した実験用検索システム以外の検索エンジン（例：Google 検索）を意図

*5 Lancers.jp: <http://www.lancers.jp>

*6 「証拠」は「根拠」、「仕組み」は「原理」と類似すると判断した。

表 2 批判的思考態度を持った人物を連想すると評価された語の例 (括弧内の数字はワーカが付与した評価値の平均値). 下線を引いた語をクエリプライミングに用いた

Table 2 Examples prime term candidates associated with critical thinkers (numbers in parentheses are the mean crowd workers ratings). Underlined terms were used for query priming in the user study.

批判的思考態度の要素	プライム語候補 Top 5				
論理的思考の自覚	<u>原理</u> (1.37)	<u>証拠</u> (1.22)	仕組み (1.22)	過程 (1.20)	証明 (1.18)
探究心	<u>調査</u> (1.37)	<u>研究</u> (1.33)	<u>検証</u> (1.31)	追求 (1.16)	比較 (0.961)
客観性	<u>比較</u> (1.27)	<u>統計</u> (1.24)	<u>分析</u> (0.980)	違い (0.745)	口コミ (0.686)
証拠の重視	<u>根拠</u> (1.74)	<u>実証</u> (1.74)	<u>データ</u> (1.56)	裏付け (1.52)	証明 (1.48)

表 3 検索タスクで用いた質問
Table 3 Search task questions.

種類	質問
Open	電球を発明した人は誰でしょうか?
	望遠鏡を発明した人は誰でしょうか?
	蒸気機関を発明した人は誰でしょうか?
	地球温暖化の原因は何でしょうか?
Closed	恐竜が絶滅した理由は何でしょうか?
	シナモンは糖尿病の症状改善に有効でしょうか?
	ビタミンCは肺炎の予防に有効でしょうか?
	イチヨウの葉は耳鳴り症状の改善に有効でしょうか?
	ココアは血圧低下に有効でしょうか?
	にんにくは風邪の予防や症状改善に有効でしょうか?

表 4 被験者の割当て
Table 4 Participants allocation.

SUI	学歴	
	大卒経験者	非大卒経験者
control	29	31
priming	26	32

せず使ってしまったため、分析の対象外とした。最終的に 118 名の被験者のデータを分析用に用いた。なお、118 名のうち 55 名は大学卒業経験者であった。

実験終了後、被験者に普段利用している検索エンジンの習熟度について、5 段階のリッカート尺度で回答してもらった (-2:まったく慣れていない, +2:非常に慣れている)。その結果、ほとんどの被験者が検索エンジンの利用方法についてある程度習熟していることを確認した (mean: 1.36, SD: 0.73)。各被験者には実験参加の報酬として 400 円を支払った。

5.2 タスク

本ユーザ実験では、各被験者は計 10 件の検索タスクに取り組んだ。表 3 に各検索タスクで用いた質問を記す。表で示したように、本ユーザ実験では、**open 質問**と **closed 質問**の 2 種類を用意した。Open 質問タスクでは、複数の解候補が考えられる質問に対する解を指定された検索システムを用いて探すタスクである。被験者には質問に対する解の候補を検索し、複数の解の候補から最終的に解を 1 つ回答するよう依頼した。Closed 質問タスクでは、指定された検索システムを用いて、与えられた質問に対する回答を *yes* か *no* のいずれかで答えるタスクである。本実験では、批判的に情報を検証することが求められる科学・技術カテゴリを取り上げ、あらかじめ学術的に正しい回答が明らか

になっている質問を発明、科学、医学カテゴリから収集し、検索タスクとして設定した。なお、質問はウェブ上に回答候補が複数存在するものを選んだ。

各検索タスクでは、被験者には用意された検索システムを用いてウェブ検索を行い、回答の候補および回答の決め手になるウェブページを探すよう依頼した。

各タスクを開始する前に、被験者には質問内容に関する事前知識を問うアンケートを行った。被験者は事前知識の程度を 5 段階のリッカート尺度で回答した (-2:まったく知識がない, +2:かなり知識がある)。結果、事前知識がないという質問 (尺度上 0 未満の回答) は被験者 1 人につき平均 8.08 個であった (SD = 1.58)。また、10 個の質問に対する事前知識の平均スコアは -1.10 であった (SD = 0.50)。このアンケート結果から、被験者の大半は質問に対する回答を事前には知らなかった、あるいは回答が思い浮かんでもあまり自信がなかったことが確認された。

5.3 実験デザインと手順

本ユーザ実験は、学歴および検索ユーザインタフェース (以下、SUI) を要因とする 2 要因 (2x2 水準) の被験者間計画で実施した。学歴要因には **大卒経験者** (大学院卒も含む) と **非大卒経験者** の 2 水準を設定した。SUI 要因には、一般的なウェブ検索エンジンで提供されているクエリ補完・推薦機能を実装したウェブ検索システム **control** とクエリ補完・推薦でクエリプライム語を提示するウェブ検索システム **priming** の 2 水準を設定した。各 SUI の詳細な説明は 5.4 節に記す。

各被験者は 2 種類ある SUI のいずれかに無作為に割り当てられた。結果として、各水準への割当て人数は表 4 のとおりとなった。被験者は Lancers.jp のサイトで実験参加

への同意確認を行った後、著者らが用意したユーザ実験用ウェブサイトに移動した。その後、被験者は4フェーズからなるユーザ実験を開始した。

第1フェーズとして、練習フェーズを用意した。このフェーズでは、被験者にタスク内容や実験用検索システムに慣れてもらうために、練習用の検索タスクに取り組むよう依頼した。

第2フェーズとして、介入フェーズを用意した。介入フェーズでは、各被験者はあらかじめ割り当てた実験用システムを使用して8件の検索タスクに取り組んだ。各検索タスクの冒頭では、以下のような導入文を提示した。

シナモンは糖尿病の症状改善に有効でしょうか？
「検索を開始する」ボタンをクリックし、実験用検索システムを使って質問に対する回答を探してください。満足する回答を見つけたら、このページに戻り、見つけた回答と回答の決め手になったウェブページのURLを報告してください。回答の決め手となったウェブページは複数報告していただいても問題ありません。

なお、本タスクには時間制限を設けなかった。被験者はタスク質問に対して十分に満足のいく回答が見つかり次第検索をやめ、回答と回答の決め手となったウェブページのURLを報告した。介入フェーズでは、open質問から4題、closed質問から4題の質問をランダムに選択し、被験者に割り当てた。提示する質問の順序は被験者ごとにランダム化した。

第3フェーズである事後フェーズは、クエリプライミングの持続性を分析するためのフェーズである。事後フェーズでは、被験者は介入フェーズで取り組んでいない2件の質問(open質問1件、closed質問1件)について、クエリ補完・推薦機能を停止した検索システムを用いて検索タスクに取り組んだ。タスクをはじめる前には、介入フェーズと同様のタスク導入文を被験者に提示した。

実験終了後、第4フェーズとして、被験者は事後アンケートに回答した。アンケートは以下の2種類の質問から構成された：(1) 実験で用いた検索システムの有効性を問う質問、(2) 検索タスクの回答を決定する際に重視した観点を問う質問。

システムの有効性については、以下の項目に関して5段階のリッカート尺度で回答するよう被験者に依頼した(-2:まったくそう思わない, +2:そう思う)。

クエリ修正を行う際のクエリ推薦・補完機能の有効性

検索システムに入力したクエリを修正するうえで、クエリ補完・推薦機能が提示した情報がどの程度有用であったか。

意志決定の観点を考える際のクエリ推薦・補完機能の有効性 ウェブページや回答候補を評価する観点を考えるうえで、クエリ補完・推薦機能が提示した情報がどの程度

有用であったか。

提示された検索結果リストの有効性 タスク回答や回答の決め手を探すうえで、提示された検索結果はどの程度有用であったか。

タスク回答を決定する際の観点については、検索タスク中に以下の項目についてどの程度重視したかを5段階で評価してもらうよう被験者に依頼した(-2:ほとんど重視しなかった, 2:非常に重視した)。

コンテンツの網羅性 コンテンツに十分な量の情報が含まれているか。

コンテンツの鮮度 コンテンツはどの程度新しい内容か。

コンテンツの客観性 コンテンツはどの程度客観的で偏りが少ないか。

コンテンツの典型性 類似する内容が書かれているウェブページが他にどの程度存在するか。

コンテンツの社会的評価・評判 ウェブページに対する社会的評価はどの程度か。

コンテンツ作成者 ウェブページのコンテンツを作成した人物がどんな人物であるか。

5.4 実験用検索システム

本ユーザ実験用に、GoogleやYahoo!といった一般的なウェブ検索エンジンに類似するSUIを持つシンプルなウェブ検索システムを開発した。この検索システムはトップページ画面と検索結果一覧ページ(SERP)画面から構成される。

実験システムでは、トップページ画面もしくはSERP画面に備えつけられた検索ボックスにクエリが入力されると、クエリ補完機能として検索ボックス直下にクエリ候補が複数表示されるようにした。SUIをcontrol条件に設定した場合、検索システムはGoogle Suggest API^{*7}を利用して、クエリ補完時にクエリに関連するクエリを最大10個提示した。SUIをpriming条件に設定した場合、システムはクエリ補完時にcontrol条件で提示していた関連クエリの半数を別のクエリで置き換えたものを表示した。置き換え後のクエリは、表2に掲載したプライム語の末尾にユーザが入力したクエリ語を付け足したもの(以下、クエリプライム)を用いた(図2(1))。なお、置換に用いるプライム語は表2で下線を引いた語からランダムに選択した。

SERP画面では、システムはMicrosoft Bing Web Search API^{*8}を用いて、入力されたクエリに関連するウェブ検索結果の一覧を提示した。提示される検索結果数は150件に固定した。各検索結果の構成要素はタイトル、スニペット、URLとした。被験者が各検索結果のタイトルをクリックすると、システムはクリックされたタイトルのウェブ

^{*7} <http://suggestqueries.google.com/complete/search>

^{*8} <https://azure.microsoft.com/services/cognitive-services/bing-web-search-api/>

ページをブラウザの別タブで提示した。SERP 画面ではクエリ推薦機能として、入力されたクエリに関連するクエリを検索結果一覧の直前と直後のエリアに提示するようにした。SUI が control 条件に設定されていた場合、システムは Microsoft Bing Web Search API を用いて関連クエリを最大 8 件推薦するようにした。SUI が priming 条件に設定されていた場合、クエリ補完機能と同様に、システムは control 条件で提示する関連クエリ語の半数をクエリプライムで置き換えたものを推薦するようにした (図 2(2))。なお、置換に用いるプライム語は表 2 で下線を引いた語からランダムに選択した。

6. 結果

118 名の被験者から 1,180 件の検索タスクの回答および検索・閲覧行動データを収集した。それらデータを解析し、SUI 要因および学歴要因が被験者の検索・閲覧行動やタスク中の意思決定に与える影響について分析した。分析は介入フェーズおよび事後フェーズに分けて行った。さらに、事後アンケート結果の分析を通じて、各種要因が被験者の検索・閲覧態度に与える影響について分析を行った。なお、収集データに正規性が確認できなかったため、整列ランク変換を用いたノンパラメトリック分散分析を行った [24]。分散分析には ARTool 統計分析パッケージを用いた*9。

6.1 タスクパフォーマンス

RQ1 に回答するために、各検索タスクの所要時間について分析を行った。なお、タスク所要時間にはタスク開始時に提示される導入文を読む時間、タスクへの回答および回答の決め手となったウェブページの URL を提出する時間も含まれる。

我々は、クエリプライミングがウェブ検索プロセスにおける批判的思考を活性化させるならば、クエリプライミングが実装された SUI (priming 条件) を用いた被験者はタスクにかかる所要時間が長くなると予想していた。しかし、表 5 と表 6 が示しているように、介入フェーズ、事後フェーズともに、タスク所要時間に関して、SUI 要因および学歴要因に統計的有意差は見られなかった。

6.2 検索・閲覧行動

RQ1, RQ3, RQ4 に答えるために、各タスクにおける被験者の検索・閲覧行動を分析した。分析した項目は以下のとおりである：

クエリ発行数 各タスク中に被験者が発行したクエリの数 (検索頻度)

SERP 訪問数 各タスク中に被験者が検索結果一覧ページ (SERP) を訪問した回数

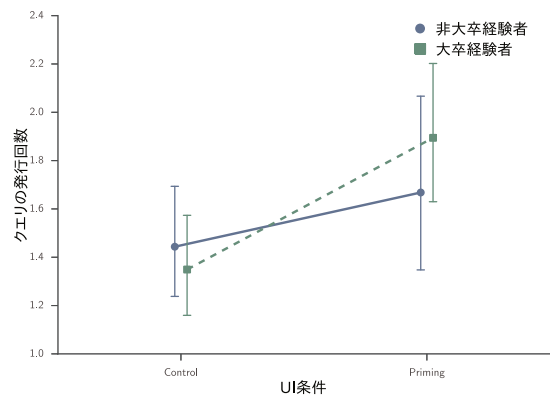


図 3 学歴と SUI ごとにみる、介入フェーズにおけるタスク中のクエリ発行回数 (エラーバーは 95%信頼区間を表す)

Fig. 3 Number of issued queries in each task during the intervention phase divided by educational background and UI condition (error bar means 95% confidence interval).

SERP を除くページの訪問数 各タスク中に被験者が SERP 以外のウェブページを訪問した回数

介入フェーズにおいて、クエリ発行回数に関しては、SUI 要因および SUI・学歴間の交互作用に統計的有意差があることが確認された (SUI: $F_{(1,114)} = 15.8, p < 0.001, \text{partial } \eta^2 = 1.22e-1$; 交互作用: $F_{(1,114)} = 10.2, p < 0.01, \text{partial } \eta^2 = 8.19e-2$)。図 3 が示すように、被験者が大卒以上の学歴を有するか否かによらず、priming 条件を用いた被験者は control 条件を用いた被験者よりも平均クエリ発行回数が多かった (大卒経験者グループの mean: 1.35 vs. 1.89 (表 5); 非大卒経験者グループの mean: 1.44 vs. 1.67 (表 5))。SUI 要因と学歴要因間の交互作用を検証するために、Mann-Whitney の U 検定を用いた単純主効果分析を行った。その結果、大卒経験のある被験者のクエリ発行回数については、priming 条件を用いた方が control 条件を用いるよりも有意に多かった ($Z = -3.49, p < 0.001, r = 0.47$)。これらの結果は、介入フェーズにおけるクエリプライミングは、被験者により多くのクエリを発行させる効果があり、その効果は大卒経験のない被験者よりも大卒経験のある被験者の方が大きかったことを示唆している。

事後フェーズについては、表 6 が示すとおり、priming 条件を用いた被験者は control 条件を用いた被験者よりも多くのクエリを発行していた (大卒経験者グループの mean: 1.24 vs. 1.88; 非大卒経験者グループの mean: 1.31 vs. 1.45)。さらに、分散分析の結果、クエリ発行回数に関しては学歴要因、SUI 要因に統計的有意差が確認された (学歴: $F_{(1,114)} = 4.71, p < 0.05, \text{partial } \eta^2 = 4.57e-2$; SUI: $F_{(1,114)} = 5.46, p < 0.05, \text{partial } \eta^2 = 3.97e-2$)。なお、学歴・SUI 条件間の交互作用には統計的有意差は確認されなかった。これらの結果は、介入フェーズで priming 条件が実装された検索 UI を用いた被験者は、その UI の使用をやめた後でも、control 条件が実装された UI を用いた

*9 <http://depts.washington.edu/madlab/proj/art/>

表 5 学歴と SUI ごとにみる，介入フェーズにおける各被験者の行動データおよびタスクの回答に関する統計 (***: 有意水準 0.001, **: 0.01, *: 0.05, ∙: 0.1)

Table 5 Participant behaviors and submitted answers during each task in the *intervention* phase broken down by educational background and UI condition (***: significance level at 0.001, **: 0.01, *: 0.05, and ∙: 0.1).

指標	大卒経験者		非大卒経験者		p-value			
	Control	Priming	Control	Priming	学歴	SUI	交互作用	
タスクパフォーマンス 所要時間 (秒)	226.8	239.7	183.4	210.2	0.48	0.48	0.54	
検索・閲覧行動	クエリ発行回数	1.35	1.89	1.44	1.67	∙	***	**
	SERP 訪問数	4.72	6.13	4.14	4.90	0.19	*	0.30
	SERP を除くウェブページ訪問数	6.06	8.13	5.90	5.97	0.36	0.28	0.40
タスク回答時に提出された根拠ページ	根拠ページ (URL) の数	1.52	1.90	1.47	1.40	*	0.19	*
	参照情報が掲載されたページ提出割合 (%)	49.6	61.1	48.4	53.9	0.41	*	0.41
	コンテンツ作成者の実在性が確認できるページ提出割合 (%)	45.7	50.4	33.0	36.3	***	0.12	0.90
	コンテンツ作成者の専門性が確認できるページ提出割合 (%)	20.2	22.7	29.3	35.1	***	0.24	1.00
信頼できるトップレベルドメインを持つページ提出割合 (%)	8.6	11.1	4.0	3.5	***	0.62	0.30	

表 6 学歴と SUI ごとにみる，事後フェーズにおける各被験者の行動データおよびタスクの回答に関する統計 (***: 有意水準 0.001, **: 0.01, *: 0.05, ∙: 0.1)

Table 6 Participant behaviors and submitted answers during each task in the *plain* phase broken down by educational background and UI condition (***: Significance level at 0.001, **: 0.01, *: 0.05, and ∙: 0.1).

指標	大卒経験者		非大卒経験者		p-value			
	Control	Priming	Control	Priming	学歴	SUI	交互作用	
タスクパフォーマンス 所要時間 (秒)	171.8	228.1	163.8	200.6	0.66	0.12	0.69	
検索・閲覧行動	クエリ発行回数	1.24	1.88	1.31	1.45	*	*	∙
	SERP 訪問数	4.30	5.98	3.60	4.89	0.11	*	0.35
	SERP を除くウェブページ訪問数	4.95	8.17	5.37	6.70	0.54	0.13	0.37
タスク回答時に提出された根拠ページ	根拠ページ (URL) の数	1.55	1.98	1.50	1.44	*	*	*
	参照情報が掲載されたページ提出割合 (%)	48.3	63.5	53.2	62.5	0.77	0.13	0.54
	コンテンツ作成者の実在性が確認できるページ提出割合 (%)	41.4	46.2	43.5	45.3	0.89	0.60	0.79
	コンテンツ作成者の専門性が確認できるページ提出割合 (%)	19.0	32.7	25.8	37.5	0.51	*	0.89
信頼できるトップレベルドメインを持つページ提出割合 (%)	3.4	15.4	8.1	7.8	0.17	*	∙	

被験者よりもクエリ発行回数が多くなったということを示唆している。

SERP 訪問数に関しては，介入フェーズにおいて，SUI 要因に統計的有意差があることが確認された ($F_{(1,114)} = 4.56$, $p < 0.05$, partial $\eta^2 = 3.84e-2$)。表 5 が示すように，学歴によらず priming 条件を用いた被験者は control 条件を用いた被験者よりも頻繁に SERP を訪問していた (大卒経験者グループの mean: 4.72 vs. 6.13; 非大卒経験者グループの mean: 4.14 vs. 4.90)。これらの結果は，priming 条件に設定された検索 UI を用いることによって，被験者はクエリ発行後もしくはウェブページ閲覧後に検索結果一覧を (再) 訪問する頻度が増えたことを示唆している。

同じ傾向は事後フェーズでも確認された ($F_{(1,114)} = 4.96$, $p < 0.05$, partial $\eta^2 = 4.17e-2$) (大卒経験者グループの mean: 4.30 vs. 5.98 (表 6); 非大卒経験者グループの mean: 3.60 vs. 4.89 (表 6))。これらの結果から，control

条件と比較して priming 条件が実装された検索 UI は，その検索 UI を使用した後でも，SERP への訪問回数を増加させる効果があったことが予想される。

SERP 以外のウェブページの訪問回数については，介入フェーズでも事後フェーズでも学歴要因および SUI 要因に統計的な有意差は確認されなかった。

以上の結果から，priming 条件が実装された検索 UI は，被験者の検索エンジンと関連した検索行動に大きな影響を与えていたと考えられる。一方で，検索エンジンが提供するページ外 (トップページ，SERP 以外のページ外) における閲覧行動については，priming 条件と control 条件間に統計的有意差はなかったと考えられる。

6.3 意志決定に用いた情報ソース

RQ2, RQ3 および RQ4 に答えるために，各被験者がタスクへの回答の根拠として提出したウェブページの分析を

行った。この分析では、提出された計 309 件のウェブページを手作業で分類し、下記の観点からクエリプライミングの証拠に基づく意志決定の促進効果について検証を行った。
根拠ページの数 各タスクで回答の根拠として提出されたウェブページの数。

コンテンツ作成者の専門性が確認できるページの提出割合
 コンテンツ作成者の専門性が確認できる記述が含まれるウェブページが提出されていた割合*10。

コンテンツ作成者の実在性が確認できるページの提出割合
 コンテンツ作成者が実在することが確認できる記述が含まれるウェブページが提出されていたタスクの割合*11。

信頼できるトップレベルドメインを持つページの提出割合
 政府系組織や学術機関を示すドメイン上で公開されているウェブページが提出されたタスクの割合*12。

参照情報が掲載されたページの提出割合 根拠として提出されたウェブページのうち、タスク回答を支持する適切な参照情報を引用しているウェブページの割合*13。

各タスクで提出された根拠ページの数については、介入フェーズにおいて学歴要因および学歴・SUI 要因の交互作用に統計的有意差があることが確認された (学歴: $F_{(1,114)} = 5.45, p < 0.05, \text{partial } \eta^2 = 4.57\text{e-}2$; 交互作用: $F_{(1,114)} = 4.04, p < 0.05, \text{partial } \eta^2 = 3.42\text{e-}2$)。図 4 に示したように、被験者が大卒経験者であった場合、control 条件を使用した被験者よりも priming 条件を用いた被験者の方がより多くの証拠ウェブページを提出した ($mean: 1.90$ vs. 1.52 (表 5))。一方、被験者が非大卒経験者であった場合、priming 条件を用いた被験者も control 条件を用いた被験者も証拠ウェブページの提出数はほとんど同数であった ($mean: 1.40$ vs. 1.47 (表 5))。Mann-Whitney の U 検定による単純主効果分析を行ったところ、大卒経験のある被験者において SUI 要因については有意傾向であった ($Z = -1.68, p = 0.094 < .1, r = 0.23$)。これらの結果から、統計的有意差は確認されなかったが、大卒経験のある被験者が priming 条件が実装された検索 UI を用いた場合、より多くの証拠ページを見つけようとした可能性が考えられる。一方、大卒経験のない被験者に対するクエリプライミングの効果は確認できなかった。

*10 本実験では、コンテンツ作成者の専門性を示す公的機関発行の認定書 (例: 医者, 弁護士資格), あるいはコンテンツ作成者の所属機関が専門性を有する組織 (政府関連機関, 学術組織, 一般に専門性を有していると考えられている株式会社) であることがウェブページ中に掲載されていた場合、コンテンツ作成者の専門性を判断可能な情報が掲載されていると定義した。

*11 ウェブページ中にコンテンツ作成者の実名や実在する組織に所属していることが明記されていた場合、コンテンツ作成者が実在することが確認できる情報が掲載されていると定義した。

*12 具体的には次のドメインを信頼できるドメインとした: *go.jp, .gov, ac.jp, and .edu*。

*13 学術機関や政府系機関が発行している情報をページ内で参照している場合、そのウェブページは参照可能な情報を掲載していると見なした。

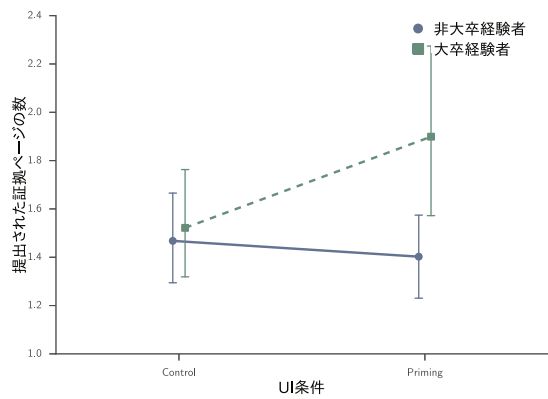


図 4 学歴, SUI 別にみる, 介入フェーズのタスク中に提出された証拠ページの数 (エラーバーは 95%信頼区間を意味する)

Fig. 4 Number of URLs posted as answer evidence in each task during the intervention phase divided by educational background and UI condition (error bar means confidence interval at 95% level).

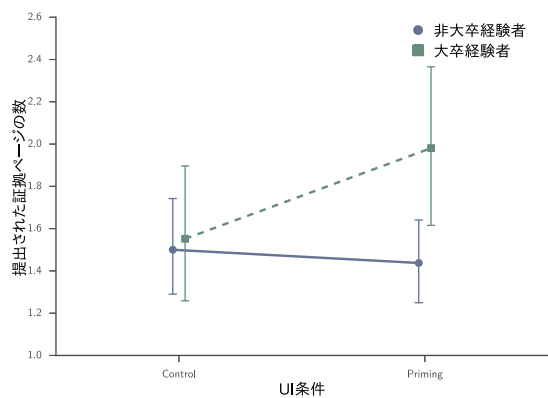


図 5 学歴, SUI 別にみる, 事後フェーズのタスク中に提出された証拠ページの数 (エラーバーは 95%信頼区間を意味する)

Fig. 5 Number of URLs posted as answer evidence in each task during the plain phase divided by educational background and UI condition (error bar means confidence interval at 95% level).

証拠ページの提出数に関する同様の傾向は、事後フェーズでも確認された。事後フェーズでは、証拠ページの提出数に対して、学歴要因, SUI 要因, 学歴・SUI 間の交互作用に統計的有意差があることが確認された (学歴: $F_{(1,114)} = 4.16, p < 0.05, \text{partial } \eta^2 = 3.52\text{e-}2$; SUI: $F_{(1,114)} = 3.99, p < 0.05, \text{partial } \eta^2 = 3.38\text{e-}2$; 交互作用: $F_{(1,114)} = 5.33, p < 0.05, \text{partial } \eta^2 = 4.47\text{e-}2$)。図 5 が示しているように、被験者が大卒経験者であった場合、priming 条件を実装した検索 UI の使用を停止した後のタスクで提出された証拠ページの数、control 条件を実装した UI を使い続けた場合に提出された証拠ページの数よりも平均的には多かった (大卒経験者グループの $mean: 1.98$ vs. 1.55 ; 非大卒経験者グループの $mean: 1.44$ vs. 1.50)。Mann-Whitney の U 検定を用いた単純主効果分析を行ったところ、大卒経験のある被験者において SUI 要因について有意傾向であるこ

とが確認された ($Z = -1.93, p = 0.054 < .1, r = 0.26$). これらの結果から, priming 条件を実装した検索 UI は, それを使用している間も使用をやめた後でも被験者により多くの証拠となるウェブページを見つけようとさせる傾向にあったことが伺える.

参照情報が掲載されたウェブページが提出された割合については, 介入フェーズにおいて SUI 要因にのみ統計的有意差が確認された ($F_{(1,114)} = 4.08, p < 0.05, \text{partial } \eta^2 = 3.46e-2$). 表 5 が示すとおり, 大卒経験のない被験者が priming 条件を用いた場合, control 条件を用いた場合と比べて, 参照情報が掲載されたウェブページを証拠情報として提出するケースが 5.5% 多かった. 被験者が大卒経験者であった場合, priming 条件を実装した検索 UI を用いた被験者は, control 条件を実装した UI を用いた被験者と比べて, 参照情報が掲載されたウェブページを証拠情報として提出するケースが 11.5% 多かった (表 5). 一方, 事後フェーズでは, 学歴要因, SUI 要因ともに, 参照情報が掲載されたウェブページの提出割合について統計的有意差は確認されなかった.

コンテンツ作成者の専門性が確認できるページが提出された割合, 信頼できるトップレベルドメインを持つページが提出された割合については, 介入フェーズでは SUI 要因に統計的有意差は確認できなかった. しかし, 学歴要因に統計的有意差があることは確認された (コンテンツ作成者の専門性が確認できるページが提出された割合: $F_{(1,114)} = 12.4, p < 0.001, \text{partial } \eta^2 = 9.82e-2$; 信頼できるトップレベルドメインを持つページが提出された割合: $F_{(1,114)} = 11.1, p < 0.001, \text{partial } \eta^2 = 8.87e-2$). 一方, 事後フェーズにおいては, SUI 要因に統計的有意差が確認された (コンテンツ作成者の専門性が確認できるページが提出された割合: $F_{(1,114)} = 4.16, p < 0.05,$

$\text{partial } \eta^2 = 3.52e-2$; 信頼できるトップレベルドメインを持つページが提出された割合: $F_{(1,114)} = 4.61, p < 0.05, \text{partial } \eta^2 = 3.89e-2$). これらの結果から, priming 条件が実装された検索 UI を使用している間は control 条件 UI と見られなかったが, その UI の使用をやめた後に関しては, priming 条件 UI を使用していた被験者は, 通常の control 条件 UI を用いていた被験者よりも信頼できるトップレベルドメインを持つページやコンテンツ作成者の専門性が確認できるページを証拠情報として提出していたといえる. ただし, 今回の実験データからは, 被験者がページ閲覧中に実際にコンテンツ作成者の専門性やトップレベルドメインに注意を払っていたかについては明確には言及できないことには留意が必要である.

コンテンツ作成者の実在性が確認できるウェブページの提出割合については, 介入フェーズでも事後フェーズでも SUI 要因について統計的有意差を確認することはできなかった.

6.4 事後アンケート分析

クエリプライミング効果の質的分析を行うために, 5.3 節で述べた事後アンケートに対する被験者の回答を分析した. 表 7 は, 各 SUI の有効性および意志決定の各種観点の重要性についての平均点を示している (-2: まったくそう思わない ~ +2: そう思う).

入力したクエリを修正するうえでのクエリ補完・推薦機能の有効性, 提示された検索結果リストの有効性については, 学歴要因, SUI 要因ともに統計的有意差は確認されなかった. すなわち, control 条件が実装された検索 UI と比較しても, priming 条件が実装された UI について被験者は特に不満を持たなかったことが伺える. 一方, 意志決定の観点を考えるうえでのクエリ補完・推薦機能の有用性

表 7 学歴および SUI 別にみる, 検索 UI の有効性および意志決定の各種観点の重要性スコア (***: 有意水準 0.001, **: 0.01, *: 0.05, .: 0.1)

Table 7 System usefulness and viewpoint for decision making in each task broken down by educational background and query completion/suggestion type (***: significance level at 0.001, **: 0.01, *: 0.05, and .: 0.1).

指標	大卒経験者		非大卒経験者		p-value		
	Control	Priming	Control	Priming	学歴	SUI	交互作用
検索 UI の有効性							
クエリ修正を行ううえでのクエリ推薦・補完機能の有効性	1.03	1.08	0.87	1.00	0.61	0.48	0.65
意志決定の観点を考えるうえでのクエリ推薦・補完機能の有効性	0.72	0.88	0.68	0.94	0.61	**	0.92
提示された検索結果リストの有効性	0.93	0.88	1.06	1.16	0.25	0.85	0.73
意志決定の観点							
コンテンツの網羅性	0.90	1.15	0.84	0.88	0.59	0.49	0.91
コンテンツの鮮度	0.24	0.04	0.35	-0.19	0.81	.	0.41
コンテンツの客観性	1.34	1.35	1.03	1.03	.	.	0.74
コンテンツの典型性	1.00	1.15	1.00	1.00	0.56	0.65	0.52
コンテンツの社会的評価・評判	-0.24	-0.27	-0.29	-0.41	0.93	0.87	0.86
コンテンツ作成者	0.34	1.11	0.16	-0.25	**	0.84	*

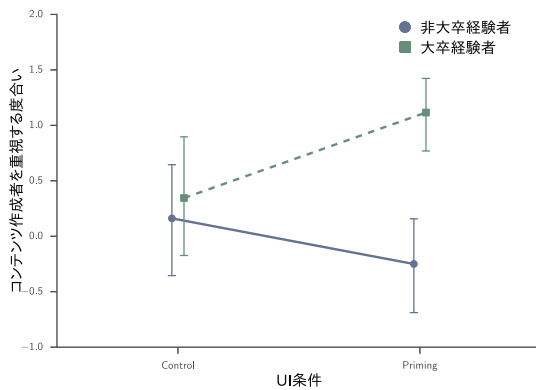


図 6 学歴, SUI 別にみる, 検索タスクの意志決定においてコンテンツ作成者を重視する度合い (エラーバーは 95%信頼区間を意味する)

Fig. 6 Extent to which participants considered content author in the search task divided by educational background and UI condition (error bar means 95% confidence interval).

については, SUI 間に統計的な有意差があることが確認された ($F_{(1,114)} = 7.88, p < 0.01, \text{partial } \eta^2 = 6.47\text{e-}2$). 表 7 が示しているように, この項目に関しては, 学歴によらず priming 条件 UI を用いた被験者は control 条件 UI を用いた被験者よりも高い得点をつけた (大卒経験者グループの $mean : 0.88$ vs. 0.72 ; 非大卒経験者グループの $mean : 0.94$ vs. 0.68). この結果から, 被験者は priming 条件を実装した検索 UI がクエリ補完・推薦時に提示した語は control 条件 UI が提示した語よりも, タスクの回答をウェブ検索で調べる際に有用であると考えた可能性がある.

検索タスクの意志決定時に重視した観点については, コンテンツ作成者の観点のみ, 学歴および学歴・SUI 間の交互作用に統計的有意差が確認された (学歴: $F_{(1,114)} = 9.93, p < 0.01, \text{partial } \eta^2 = 8.01\text{e-}2$; 交互作用: $F_{(1,114)} = 5.63, p < 0.05, \text{partial } \eta^2 = 4.70\text{e-}2$). 図 6 が示しているように, 被験者が大卒経験者であった場合, priming 条件を実装した検索 UI を用いた被験者は control 条件 UI を用いた被験者よりも, コンテンツ作成者の重要度が平均的に高かった ($mean : 1.11$ vs. 0.34 (表 7)). また, 被験者が大学を卒業していない場合, priming 条件 UI を用いた被験者は control 条件 UI を用いた被験者よりも, コンテンツ作成者に関する平均重要度が小さかった ($mean : -0.25$ vs. 0.16 (表 7)). Mann-Whitney の U 検定を用いた単純主効果分析を行った結果, 大卒経験のある被験者において SUI 要因は有意傾向であることが分かった ($Z = -1.78, p = 0.074 < .1, r = 0.24$). 一方, 大卒経験の被験者において SUI 要因に統計的有意差は確認されなかった. これらの結果から, 大卒経験のある被験者に対しては, クエリプライミングはコンテンツ作成者に対する注意意識を高めた可能性が考えられる.

7. 考察

7.1 仮説の検証

本節では, ユーザ実験の結果をふまえて 3.3 節で設定した RQ1~4 に対する回答について議論を行う.

RQ1 と RQ4 を検討するために, 大学卒業経験がある被験者となない被験者のタスク結果および検索行動を比較した. タスク中の行動データを分析ところ, 学歴, SUI 条件の違いがタスク所要時間に影響するという事は認められなかった (6.1 節). さらに, ページ閲覧行動を分析した結果, ウェブページの閲覧回数に対するクエリプライミングの影響も見受けられなかった (6.1 節). クエリプライミングが注意深い情報探索を促進すると仮定した場合, クエリプライミングが実装された UI を用いた被験者は, 慎重な意志決定を行うために検索タスクにより時間をかけるようになったり, より多くのウェブページを閲覧するようになったりすると予想していた. しかし, 実験結果からはこの予想を支持する結果は得られなかった.

一方, 検索行動を分析した結果, クエリプライミングが実装された UI は, クエリの発行頻度や検索結果一覧ページへの再訪問回数を増加させる傾向が明らかになった (6.2 節). 特に, クエリプライミングは大卒経験のある被験者のクエリ発行頻度に強い影響があることが明らかになった. 事後アンケートの分析によると, クエリプライミングを実装した検索 UI を用いた被験者は通常の検索インタフェースを用いた被験者と比べても, (クエリプライミングが行われる) クエリ補完・クエリ推薦機能に特に不満を感じた様子はなかった (6.4 節). また, アンケート結果は, クエリ発行回数が多くなった理由がクエリ補完・クエリ推薦機能のパフォーマンスの低さによるものでないことも示唆している. このことから, クエリプライミングは限られた時間の中でどのウェブページを閲覧すべきかを注意深く選択するために, クエリの発行・修正頻度や検索結果一覧ページの (再) 訪問回数を増加させる効果があったと思われる.

RQ2 と RQ4 について検討するために, 被験者がタスク中に意志決定の証拠情報として提出したウェブページの URL および事後アンケート結果を分析した. 事後アンケートを分析した結果, 被験者のコンテンツ作成者に対する意識にクエリプライミングが何らかの影響を与えていることが分かった (6.4 節). 図 6 が示しているように, 大卒経験者がクエリプライミングが実装された検索 UI を用いた場合, コンテンツ作成者に対する意識が高まったのに対して, 大学を卒業していない被験者は同じ検索 UI を用いるとコンテンツ作成者に対する意識は低くなった.

しかし, タスク中に提出された意志決定の証拠ページ URL を分析したところ, 異なる解釈が得られた. 表 5 によると, クエリプライミングを実装した検索 UI の使用中は, コンテンツ作成者の実在性, コンテンツ作成者の専門

性、トップレベルドメインなど、コンテンツ作成者に関する観点による証拠ページの決定が促進されるという統計的事実は得られなかった。一方、表 5 が示すように、クエリプライミングが実装された検索 UI を用いた被験者はそれを用いなかったユーザよりも参照情報が掲載されたウェブページを証拠ページとして提出することが有意に多かった。さらに、図 4 によると、クエリプライミング UI を用いた大卒経験者は、それを用いなかった大卒経験者よりも多くの証拠ページを提出する傾向にあった。このことから、大卒経験のある被験者がクエリプライミングが実装された検索 UI を用いると、本人が意識していなかったとしても、裏付け可能な参照情報が掲載されたウェブページを複数を探査するようになったと考えられる。以上の議論から、クエリプライミングは被験者の注意深い情報探索を促進し、その効果は大学の卒業経験がある被験者に強く表れたと結論づける。

RQ3 に関しては、実験結果からはクエリプライミングが実装された検索 UI の使用後もクエリプライミングの効果が持続することは示せなかった。クエリ発行回数と検索結果一覧ページの訪問数については、クエリプライミング UI の使用中も使用後も UI が被験者に影響を与えたことを確認している (表 5, 表 6 および図 3)。さらに、大卒経験のある被験者に関してのみ、クエリプライミング UI の使用後も参照情報が掲載されたウェブページの提出数を増加させる効果が持続することが確認されている (図 4, 図 5)。しかし、クエリプライミング UI を使用中は参照情報が掲載されたウェブページを証拠ページとして提出する被験者が有意に多かったのに対して、クエリプライミング UI の使用を終えた後はそのような効果は統計的には確認できなかった。回答の根拠ページを選ぶ基準としてコンテンツ作成者の専門性やトップレベルドメインが考慮されているかという点については、各 UI の使用後の比較では、クエリプライミングが実装された検索 UI と通常の実験 UI との間に統計的有意差が確認された (表 6)。しかし、クエリプライミング UI 使用時には有意差は確認されなかった (表 5)。直感的には、クエリプライミングがコンテンツ作成者の専門性やトップレベルドメインの考慮に影響を及ぼすならば、少なくともクエリプライミング UI を使用中にはその効果が現れることが期待される。しかし、実験結果はこれを支持しなかった。今回行ったユーザ実験では、介入フェーズにおけるタスクが計 8 回だったのに対し、クエリプライミングを実装した検索 UI および通常の実験 UI の使用を停止した未使用フェーズにおけるタスクはわずかに 2 回であった。それゆえ、未使用フェーズの結果が安定しなかったことも予想される。今後は、クエリプライミング効果の持続性を検証するために、実験計画を再設計し再度ユーザ実験を行う必要がある。また、今回のオンライン実験では一部の被験者が設計者側が意図していない実験シス

テムの使い方をすることがあったため、オンライン実験と並行して研究室実験の実施も検討する必要がある。

7.2 提示するプライム語の選択について

本稿で提案したクエリプライミングでは、単純化のため、クエリに非依存なプライム語を提示する方針を採用した。プライム語を設計する際に用いた検索トピックは一般消費者に馴染みのあるトピックである一方、ユーザ実験で用いた検索トピックは学術的なトピックであり、両トピックの文脈には意味的な乖離があった。しかし、実験結果から、クエリの文脈に依存しないプライム語を設計・提示することで、トピックによる違いの影響を大きく受けずに、批判的な情報探索をある程度促進できたと考える。

実験結果からクエリ非依存のプライム語でも一定のクエリプライミング効果を発生させることが確認されたが、クエリの文脈を考慮してプライム語を選択・提示した方が、より良いクエリプライミング効果を期待できると考えられる。たとえば、「ダイエット食品の購入」のように一般消費者向けの検索タスクであれば、「失敗談」や「口コミ」といったプライム語候補が考えられる。一方、「邪馬台国の所在地」のように歴史的な事実を検索するタスクであれば、「異なる学説」や「文献」といったプライム語候補がより自然でかつ有効であると考えられる。クエリプライミングの効果をより高めるためにも、今後はクエリの文脈を考慮したプライム語の発見方法およびその提示効果について検討する必要がある。

7.3 批判的思考態度と情報探索方略の関係

本稿で提案したクエリプライミングは、批判的な情報探索を行うための態度を喚起することに焦点を当てていた。ユーザ実験の結果、クエリプライミングによってある程度批判的な情報探索行動が促進されたことが確認されたが、ユーザがより効果的に批判的な情報探索行動を行うためには、態度の喚起に加えて、批判的な情報探索のための方略を伝えることも重要であると考えられる。Meola が述べているように、ウェブのような玉石混淆の情報源から正確な情報を取得するには、複数の情報を比較することや証拠情報を探すことに加え、図書館や学術機関などウェブ以外の専門的な情報源あるいは専門家をあたることも重要である [16]。それゆえ、クエリプライミングの効果を補完するためにも、検索トピックやユーザの検索状況に応じて、有効な情報探索方略を示唆するシステムなども検討する必要がある。

8. むすび

本稿では、クエリ推薦やクエリ補完機能の使用時に注意深い情報探索を促すための機能、クエリプライミングを提案した。クエリプライミングを実装した情報検索インタフェースは、批判的思考を喚起する単語をユーザに提示す

ることで、注意深くウェブページを探索することは暗黙的に促す。クラウドソーシングを用いたユーザ実験を行った結果、クエリプライミングによって被験者の検索行動が変化し、検索回数や検索結果一覧ページを再訪問する回数が増加した。また、クエリプライミングが実装された検索インタフェースを用いた被験者、特に大卒経験のある被験者は、参照情報が掲載されているウェブページに注目しウェブ探索をするようになることが確認された。

ユーザ実験の結果、クエリプライミングが実装された検索インタフェースは、それを使用しているときには一定の効果をもたらすこと確認されたが、使用を停止した後も同様の効果が持続するかについては十分な結果が得られなかった。今後は、クエリプライミング効果の持続性について焦点を当てた研究室実験を行う予定である。

情報検索の分野では、ユーザの情報要求と文書の適合性に焦点を当てた情報アクセスシステムについて多くの研究がなされている。本研究で得られた知見は、機械による自動的な情報取得・評価だけでなく、注意深い情報探索や意思決定を促進する情報検索インタラクションの設計に寄与できると期待している。

謝辞 本研究はJSPS科研費JP16H01756, JP16H02906, JP17K17832, JP18H03494, JP18KT0097の助成を受けたものです。ここに記して謝意を表します。

参考文献

- [1] Agapie, E., Golovchinsky, G. and Qvarfordt, P.: Leading People to Longer Queries, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2013*, pp.3019–3022, ACM (2013).
- [2] American Library Association (ALA): *Information literacy competency standards for higher education* (2000).
- [3] Bargh, J.A., Chen, M. and Burrows, L.: Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action, *Journal of Personality and Social Psychology*, Vol.71, No.2, pp.230–244 (1996).
- [4] Bateman, S., Teevan, J. and White, R.W.: The Search Dashboard: How Reflection and Comparison Impact Search Behavior, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2012*, pp.1785–1794, ACM (2012).
- [5] Castillo, C., Mendoza, M. and Poblete, B.: Information Credibility on Twitter, *Proc. 20th International Conference on World Wide Web, WWW 2011*, pp.675–684 (2011).
- [6] Dong, X.L., Gabrilovich, E., Murphy, K., Dang, V., Horn, W., Lugaresi, C., Sun, S. and Zhang, W.: Knowledge-based Trust: Estimating the Trustworthiness of Web Sources, *Proc. VLDB Endowment*, Vol.8, No.9, pp.938–949 (2015).
- [7] Ennals, R., Trushkowsky, B. and Agosta, J.M.: Highlighting Disputed Claims on the Web, *Proc. 19th International Conference on World Wide Web, WWW 2010*, pp.341–350 (2010).
- [8] Ennis, R.H.: A taxonomy of critical thinking dispositions and abilities, *Series of books in psychology, Teaching thinking skills: Theory and practice*, Baron, J.B. and Sternberg, R.J. (Eds.), W H Freeman/Times Books/Henry Holt & Co, New York, pp.9–26 (1987).
- [9] Harvey, M., Hauff, C. and Elsweiler, D.: Learning by Example: Training Users with High-quality Query Suggestions, *Proc. 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2015*, pp.133–142, ACM (2015).
- [10] Jeong, S., Mishra, N., Sadikov, E. and Zhang, L.: Domain Bias in Web Search, *Proc. 5th ACM International Conference on Web Search and Data Mining, WSDM 2012*, pp.413–422, ACM (2012).
- [11] Adobe Systems Incorporated: *The State of Content: Rules of Engagement for 2016* (2015).
- [12] Kahneman, D.: *Thinking, fast and slow*, Macmillan (2011).
- [13] Kelly, D., Cushing, A., Dostert, M., Niu, X. and Gyllstrom, K.: Effects of popularity and quality on the usage of query suggestions during information search, *Proc. 28th SIGCHI Conference on Human Factors in Computing Systems, CHI 2010*, pp.45–54, ACM (2010).
- [14] Leong, C.W. and Cucerzan, S.: Supporting Factual Statements with Evidence from the Web, *Proc. 21st ACM International Conference on Information and Knowledge Management, CIKM 2012*, pp.1153–1162, ACM (2012).
- [15] Liao, Q.V. and Fu, W.-T.: Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information, *Proc. 31st ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)*, pp.2359–2368, ACM (2013).
- [16] Meola, M.: Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation, *portal: Libraries and the Academy*, Vol.4, No.3, pp.331–344 (2004).
- [17] Mitra, B., Shokouhi, M., Radlinski, F. and Hofmann, K.: On user interactions with query auto-completion, *Proc. 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2014*, pp.1055–1058, ACM (2014).
- [18] Morris, M.R., Teevan, J. and Panovich, K.: What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior, *Proc. 28th ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2010*, pp.1739–1748, ACM (2010).
- [19] Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S. and Tanaka, K.: Trustworthiness Analysis of Web Search Results, *Proc. 11th European Conference on Research and Advanced Technology for Digital Libraries, ECDL 2007*, pp.38–49, Springer (2007).
- [20] Pascarella, E.T.: The development of critical thinking: Does college make a difference?, *Journal of College Student Development*, Vol.30, No.1, pp.19–26 (1989).
- [21] Pasternack, J. and Roth, D.: Latent Credibility Analysis, *Proc. 22nd International Conference on World Wide Web, WWW 2013*, pp.1009–1020 (2013).
- [22] Tsui, L.: Courses and Instruction Affecting Critical Thinking, *Research in Higher Education*, Vol.40, No.2, pp.185–200 (1999).
- [23] White, R.: Beliefs and Biases in Web Search, *Proc. 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2013*, pp.3–12, ACM (2013).
- [24] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins,

J.J.: The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2011*, pp.143-146, ACM (2011).

[25] Yamamoto, Y. and Shimada, S.: Can Disputed Topic Suggestion Enhance User Consideration of Information Credibility in Web Search?, *Proc. 27th ACM Conference on Hypertext and Social Media, HT 2016*, pp.169-177, ACM (2016).

[26] Yamamoto, Y. and Tanaka, K.: Enhancing Credibility Judgment of Web Search Results, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2011*, pp.1235-1244, ACM (2011).

[27] Yamamoto, Y. and Yamamoto, T.: Query Priming for Promoting Critical Thinking in Web Search, *Proc. 3rd ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2018)*, pp.12-21, ACM (2018).

[28] Yamamoto, Y., Yamamoto, T., Ohshima, H. and Kawakami, H.: Web Access Literacy Scale to Evaluate How Critically Users Can Browse and Search for Web Information, *Proc. 10th ACM Conference on Web Science (WebSci 2018)*, pp.97-106, ACM (2018).

[29] 平山るみ, 楠見 孝: 批判的思考態度が結論導出プロセスに及ぼす影響: 証拠評価と結論生成課題を用いての検討, *教育心理学研究*, Vol.52, No.2, pp.186-198 (2004).

付 録

A.1 収集したプライム語候補

本研究で収集したプライム語候補の一部をここに記す。なお、誌面の都合上、2名以上のワーカから回答があった語を掲載している。

- 批判的思考態度を持った人物から連想されるキーワード (表 A.1),
- および批判的思考態度を持った人物が任意のトピックについてウェブ検索を行う際、クエリ入力時に利用されると思われるキーワード (表 A.2).

表 A.1 批判的思考態度を持った人物から連想されるキーワードの一部

Table A.1 Prime term candidates associated with critical thinkers.

キーワード
冷静, 慎重, 公平, 好奇心, 根拠, 理由, データ, 冒険原因, 好奇心旺盛, 論理的, レビュー, 現実的, ランキング
口コミ, 実証, 思慮深い, 比較, 熱心, 理知的, 理論
生真面目, 証明, 人格者, 原理, 合理的, 推理, 柔軟, 検索
検証, 熟慮, 研究, 積極的, 統計, 臨機応変, 評判, 調査
頑固, 頭脳明晰, メガネ, ロジック, 仕組み, 俯瞰, 優柔不断
冷徹, 几帳面, 分析, 堅物, 客観的, 懐疑的, 旅行, 歴史
沈着, 無感情, 理屈, 要人, 画像, 疑い深い, 疑心暗鬼
真面目, 研究者, 確実性, 結果, 緻密, 考察, 興味
落ち着き, 行動, 裏付け, 証拠, 評価, 論文, 追求, 過程

表 A.2 批判的思考態度を持った人物が任意のトピックについてウェブ検索を行う際、クエリ入力時に利用されると思われるキーワードの一部

Table A.2 Query terms that critical thinkers would be likely to use when searching for a given topic.

キーワード
口コミ, 比較, 評判, おすすめ, 実績, 最新, 人気, 価格
レビュー, 機能, 評価, メリット, ランキング, リスク, 効果
効果的, 成功, 注意点, 確実, 種類, 簡単, 金額, エビデンス
コストパフォーマンス, デメリット, データ, ポイント
メーカー, 主流, 仕様, 件数, 体験談, 利益, 売れ筋, 学校
必要なもの, 感想, 手順, 料金, 方法, 書類, 流れ, 海外
得点, 生存率, 統計, 裏技, 詐欺, 論文, 講師, 費用, 速度
運動, 違い, 金利, 食事



山本 祐輔 (正会員)

静岡大学情報学部講師。2011年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。情報検索、情報の信憑性、態度・行動変容のインタラクションに関する研究に従事。ACM, 人工知能学会, 日本データベース学会, 日本デザイン学会等各会員。



山本 岳洋 (正会員)

京都大学大学院情報学研究科社会情報学専攻助教。2011年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。主に情報検索におけるユーザインタラクションやユーザ理解に関する研究に従事。日本データベース学会, ACM, 電子情報通信学会各会員。

(担当編集委員 張 建偉)