

健康情報検索におけるユーザのクエリ作成能力の調査

浜島聡一郎[†] 山本 岳洋^{††} 大島 裕明^{†,††}

[†] 兵庫県立大学 応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

^{††} 兵庫県立大学 社会情報科学部 〒651-2197 兵庫県神戸市西区学園西町 8-2-1

E-mail: [†]{aa20w509,ohshima}@ai.u-hyogo.ac.jp, ^{††}t.yamamoto@sis.u-hyogo.ac.jp

あらまし 身近な健康情報についてウェブ上で検索することは多い一方で、その情報の質は多種多様である。本研究では、現在のユーザがどの程度効果的にウェブ検索を通じて健康情報を獲得できているのかを調べるため、ユーザがウェブ検索時に作成する検索クエリに着目する。そして、ユーザが作成する検索クエリがどの程度効果的に質の高い健康情報を検索可能なのかを調査する。実験では、クラウドソーシングを用いて健康情報を検索するタスクを実験参加者に与えることで検索クエリを収集する。その後、得られた検索クエリを用いて検索結果を収集し、そのページの適合性や質を評価し、検索クエリとページの質の関係性を分析する。さらに、それらが検索専門性やeヘルスリテラシーとどのような関係にあるのかについても分析する。

キーワード 情報検索, 健康情報検索, 検索クエリ

1 はじめに

近年、多くのユーザがウェブ上で身近な健康情報について探している[2][8][11]。厚生労働省の報告では、約75%のユーザが健康に関する情報にウェブ上で触れていることが報告されている[21]。その一方で、健康情報の半数以上が専門家によってレビューが行われていないという報告もあり[18]、その質は多種多様である。そのような中、若年層のユーザの多くがウェブ上の情報を信用しているという報告や[3]、ユーザは得たい情報を獲得できたと感じたとしてもその情報が正確ではない場合が多いという報告[10]もあり、質の高い健康情報を獲得することは重要な課題であると考えられる。

ウェブ検索を用いて質の高い健康情報を獲得する上で重要なことの1つは検索エンジンに投入する検索クエリであると考えられる。多くの質の高い健康情報を検索可能な検索クエリであれば、ユーザは効率的に質の高い情報を獲得できる一方で、健康情報として質の低い、あるいは不適合な情報しか検索できない検索クエリであれば、ユーザは質の低い情報だけを基に判断をする必要となる。ここで、質の高い情報とは、JAMA ベンチマーク基準[17]に代表されるような、記事の著者が記載してあった引用元が明記されているような情報のことである。

そこで本研究では、現在のユーザがどの程度効果的にウェブ検索を通じて健康情報を獲得できているのかを検索クエリに着目して分析する。現在のユーザがどのような検索クエリを作成して健康情報を検索するのか、また、その検索結果の質はどの程度なのかを明らかにすることで、質の高い健康情報を検索するための技術に関して知見を得ることが期待される。また、ユーザの検索能力は検索エンジンに対する習熟度[19]やドメイン知識[9]に関係があると考えられる。そこで、ユーザの検索エンジンに対する習熟度や、質の高い健康情報を獲得する際に必要となるリテラシーと作成する検索クエリとの関係について

も分析する。

本研究でのリサーチクエスチョンは以下のとおりである。

RQ1 健康情報検索における、ユーザが作成した検索クエリと得られるページの質はどのような関係か。

RQ2 RQ1は検索専門性の有無によって変わるか。また、どのように変わるか。

RQ3 RQ1は健康情報に関するリテラシーの有無によって変わるか。また、どのように変わるか。

身近な健康情報についてウェブ上で検索することは多い一方で、その情報の質は多種多様である。本研究では、現在のユーザがどの程度効果的にウェブ検索を通じて健康情報を獲得できているのかを調べるため、ユーザがウェブ検索時に作成する検索クエリに着目する。そして、ユーザが作成する検索クエリがどの程度効果的に質の高い健康情報を検索可能なのかを調査する。実験では、クラウドソーシングを用いて健康情報を検索するタスクを実験参加者に与えることで検索クエリを収集する。その後、得られた検索クエリを用いて検索結果を収集し、そのページの適合性や質を評価し、検索クエリと質の関係性を分析する。さらに、それらが検索専門性やeヘルスリテラシーとどのような関係にあるのかについても分析する。

これらのリサーチクエスチョンを明らかにするため、「ココアが高血圧の改善に効果的であると主張しているページを検索する」といった4つの健康情報検索タスクを実験参加者に与えることで検索クエリを収集する実験をクラウドソーシングを用いて実施した。また、その際にアンケートを用いてユーザの検索専門性やeヘルスリテラシー[14]についても評価した。そして、収集された検索クエリを用いてウェブ検索を行い、検索クエリと得られたページの質との関係性を評価し、実験で得られた検索クエリとページの質との関係を分析した。

2 関連研究

2.1 ユーザの情報獲得能力に関する調査

ユーザが要求された情報にどの程度アクセスできるのか、その要因などについても調査が行われている。Aula らは、ラボでの実験とオンラインでの実験を行い、ユーザが発見が困難な情報に対しての行動について調査した。調査の中で見られた傾向として、ユーザはより多様なクエリ、検索演算子を用いることで情報を獲得しようとしていた [1]。

Bond らは、健康記事に関するアクセスに関して検索に慣れている人と不慣れな人で対比実験を行った。この実験においては、情報に対するアプローチの仕方次第で獲得できる情報の時間効率が変化することが報告されている [4]。

Liu らは、医療情報のドメイン知識が目的の情報獲得に与える影響について明らかにしている。ドメイン知識のあるユーザは扱うトピックに対しての情報をを用いることで、検索タスクに効率的に取り組むことができることを報告している [9]。

Pothirattanachaikul らは、健康情報に対する事前の信念の違いによって検索後の判断が影響されることを明らかにしている [15] [16]。

2.2 ウェブ上の健康情報の質

ユーザが日常的に検索を行っている健康情報の質についての調査が行われている。Kiyotaka らは、健康情報のメンタルヘルスに関する情報を収集し、記載されている情報 37 件に対して評価を行った。評価は 3 人の評価値のスコアの平均を使用した。評価指標としては、DISCERN スコア [5] を用いた。結果として、すべてのサイトにおいて情報量が不足していることが報告されている [13]。

石松らは、インターネット上に存在しているうつ病の診断に関する情報に対しての評価を行っている。評価手法としては、Silberg ら [17] の評価指標を使用している。加えて、自己評価尺度の種類、サイトを運営している種類についても評価している。自己評価尺度の種類とは、自己評価での診断では確定診断が行えないことを明記しているかを意味している。また、うつ病の診断を行うサイトにおける診断する要素、解決方法などの情報が不足しているサイトが約 7 割以上のウェブサイトある不適切な独自のうつ病評価尺度を用いていることを明らかにしている [23]。

他の健康情報として、肺がんに関する情報の質についての調査も行われている。Goto らは、日米間の肺がんに関する健康情報の国ごとの質の差を明らかにしている。日米間での肺がんに関する日本のウェブ検索では治療法の商用サイト、個人が管理しているサイトの表示が多くみられることを明らかにした。加えて、標準治療や補完医療に関する情報提供の割合の差も明らかにしている [7]。根本的な問題として、日本に大規模な健康情報に対するプラットフォームがないことが質の高い健康情報に対してアクセスすることが困難になっているという報告がされている [12]。以上のように、健康情報に関しては様々な議論が

行われている。そして、ウェブ上の健康情報の質の低さを指摘する報告が多くなされている。そのような状況の中では、ユーザ自身が質の高い健康情報を検索するための技術を備えたり、質の高い情報を検索するシステムが重要となると考えられる。

3 健康情報検索タスクを用いた検索クエリの収集

本研究ではクラウドソーシングを用いて、実験協力者に健康情報検索タスクを課し検索クエリを投入してもらう実験を行った。本節では、実験の手順、実験で用いた検索タスク、検索専門性と健康情報に関する知識を調査するためのアンケート調査について説明する。

3.1 手順

実験はクラウドソーシングを用いて実験協力者を募集した。実験を行うにあたり、調査内容を学術研究に利用することを説明し、また、説明通りにタスクが行われない場合は報酬を支払わない可能性があることを説明し、実験に参加してもらった。

実験では、まず、実験協力者に 3.2 節で述べる検索クエリ作成タスクを 4 件実施してもらい、それぞれの健康情報検索タスクに対して検索クエリを作成してもらった。その後、3.3 節および 3.4 節で述べるアンケート項目に回答してもらい、最後に、年齢や職業、学歴などのデモグラフィック属性について回答してもらった。なお、デモグラフィック属性に関する質問項目に対しては、実験協力者が回答しないという回答を行うことができるようにした。

3.2 検索クエリ作成タスク

実験タスクの健康情報は、コクラン [6] の補完医療の項目¹において、検証が行われ結果がレビューされている情報の中から 4 つを著者らが選び、検索タスクを作成した。コクラン²は、世界 130 カ国の医療に関する医者、看護師、研究者などによって運営されている団体である。選んだ健康情報は、ある症状に対しての予防法や治療法についての情報を対象としており、実験協力者にとって身近であると思われるものを著者らが選んだ。

実験にあたり、ある治療法がある症状に効果的であると主張しているページを検索する検索タスクを 2 件、ある治療法がある症状に効果的でなかったりエビデンスが不足していると主張しているページを検索する検索タスクを 2 件の計 4 件を用意した。表 1 に著者らが用意した 4 件の検索タスクを示す。たとえば、表中のタスク ID1 は、ココアの摂取が高血圧の改善に効果的であると述べているページを検索する、という検索タスクを表している。

実験にあたり、検索ユーザに対して質の高い健康情報を検索するという情報要求を持ってもらうため、次のような説明文をそれぞれのタスクごとに実験協力者に提示した。たとえば、タスク ID3 のタスクでは、『あなたの親族、知人にてんかんを患っている方がいます。あなたは、その方の症状を改善したい

1 : <https://cam.cochrane.org/cochrane-reviews-related-complementary-medicine>

2 : <https://www.cochrane.org/>

と思っています。そこで、ウェブ上でてんかんに効果的である方法を探していたところ、このページ³を見つけました。しかし、情報に対して検証が必要であると考えたあなたは、「てんかんに鍼治療が効果的ではない」と主張している発見したページのような、質の高いほかのページを探そうとしています。その際に、考えられる検索キーワードを3種類作成してください。』という文章を提示した。加えて、一般的に健康情報として質が高いと考えられるコクランのページをユーザに例示することで、単に「てんかんに鍼治療が効果的ではない」ページだけでなく、そのようなページの中でも質の高いページを探すことが求められていることを与えることができるのではないかと考えた。

上記に示した検索タスクを実施し、各タスクについて最大3種類の検索クエリを実験協力者に作成してもらった。なお、順序効果の影響を低減するため、実験では、実験協力者の半数は、ID1、ID2、ID3、ID4の順でタスクを実施してもらい、残りの半数はID3、ID4、ID1、ID2の順でタスクを実施してもらった。

3.3 検索専門性に関するアンケート

検索エンジンに関する習熟度を測定するためのアンケートを行った。本研究では、二重引用符やNOT検索といった検索エンジンの高度な機能に対する使用頻度を検索専門性と定義した。アンケート項目として山本らの研究において用いられていた6つの質問項目を本研究では使用した[22]。具体的には、以下の6つの質問項目に5段階のリッカート尺度（1. したことがない～5. 常にする）で回答してもらった。

- ウェブ検索エンジンで検索キーワードを作成する際、二重引用符(“)を用いて特定の言葉を指定する検索を使用したことはありますか。例) “厚生労働省” ニュース
- ウェブ検索エンジンで検索キーワードを入力する際、特定のキーワードを除外する、-をつけて検索を使用したことはありますか。例) 喫茶店 -渋谷
- ウェブ検索エンジンの検索結果を最終更新日時で絞り込む機能をしようしたことはありますか。例) スポーツ after:2018/09/11
- ウェブ検索エンジンの検索結果を特定サイトやドメインに絞り込む機能を使用したことはありますか。例) 食中毒 site:go.jp
- ウェブページをクリックする前にページのドメインをチェックしますか。
- 検索エンジンの検索ツールを使用して、最近作成・更新された情報を閲覧しますか。

3.4 eヘルスリテラシーに関するアンケート

質の高い健康情報を獲得するために必要な知識を測定するために、本研究では実験協力者のeヘルスリテラシーを測定した。インターネット上のヘルスリテラシーは、eヘルスリテラシーと呼ばれており、eヘルスリテラシーとは、Normanらによると[14]『インターネット上で健康情報を検索し、内容を評

価し取得した健康情報を自分の健康問題解決に向けて活用する能力』[20]と定義されている。eヘルスリテラシーを調査する質問項目として、光武らのアンケートを用いた[20]。具体的には、以下の8つの質問項目に検索専門性のアンケートと同様の5段階のリッカート尺度で回答してもらった。

- 私は、インターネット上でどのような健康情報サイトが利用できるか知っている。
- 私は、インターネット上のどこに役立つ健康情報サイトがあるか知っている。
- 私は、インターネット上で役立つ健康情報サイトの見つけ方を知っている。
- 私は、自分自身の健康状態についての疑問を解決するために、どのようにインターネットを使用すればよいか知っている。
- 私は、インターネット上で見つけた健康情報の活用方法を知っている。
- 私は、インターネット上で見つけた健康情報サイトを評価することができるスキルがある。
- 私は、インターネット上の質の高い健康情報サイトと質の低い健康情報サイトを見分けることができる。
- 私は、健康状態について判断する際に、インターネットからの情報を活用する自信はある。

3.5 実験参加者と収集した検索クエリ

実験は2020年12月29日から2021年1月5日にかけて、クラウドソーシングサービスであるランサーズ⁴を用いて100名の実験協力者からデータを収集した。100名の実験協力者には報酬をそれぞれ100円支払った。収集後のデータを分析したところ、100名中10名の実験協力者が、検索クエリを3種類入力する箇所に、それぞれ1語の検索クエリを入力していると推測されたため、これら10名のデータを分析から除去し、90名の実験協力者からのデータを分析に用いた。実験協力者の90名のデモグラフィック属性は、男性62名、女性28名、最も多い年代は40代であり37名であった。

90名の実験協力者から4つの検索タスクについてそれぞれ最大3種類の検索クエリを投入してもらい、のべ1,080件の検索クエリを収集した。分析に先立ち、検索クエリ中に含まれる半角空白と全角空白を統一し正規化を行った。最終的に、それぞれのタスクで得られた検索クエリの重複を省いた種類数を、表1に示すタスクID順にあげると、ID1では191種類、ID2では193種類、ID3では187種類、ID4では190種類となった。

4 分 析

本節では、まずクラウドソーシングで行ったアンケートに基づいた実験参加者のグループ分け、検索クエリの評価について述べた後、分析結果について述べる。

3 : <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD005062.pub4/ull/ja#CD005062-abs-0004>

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

表 1 実験で用いた健康情報検索タスク。Cochrane ID は効果の有無の判断の根拠としたコクランの記事の ID である。

タスク ID	治療法	症状	効果	コクラン ID
ID1	ココア	高血圧	効果がある	CD008893
ID2	音楽	不眠症	効果がある	CD010459
ID3	鍼治療	てんかん	効果的でない、エビデンスが不足している	CD005062
ID4	イソフラボン	高コレステロール血症	効果的でない、エビデンスが不足している	CD009518

表 2 検索専門性と e ヘルスリテラシーの度数分布表

階級	検索専門性	e ヘルスリテラシー
(0, 1]	5	0
(1.0, 2.0]	49	9
(2.0, 3.0]	31	49
(3.0, 4.0]	4	27
(4.0, 5.0]	1	5

4.1 実験参加者のグループ分け

検索専門性や e ヘルスリテラシーの違いによる作成される検索クエリの違いを分析するため、以下の方法で検索専門性と e ヘルスリテラシーに応じてそれぞれユーザを 2 グループに分割した。

検索専門性: 6 つの質問項目から得られる値の平均をその実験参加者の検索専門性のスコアとした。90 名の実験参加者の検索専門性のスコアの平均は 2.03, 標準偏差 0.69, 最小値 1.00, 最大値 4.50 であった。スコアの上位 20% の実験参加者 17 名を「検索専門性の高いユーザ」、残りの 80% にあたる 73 名を「検索専門性の高くないユーザ」とした。

e ヘルスリテラシー: 8 つの質問項目から得られる値の平均をその実験参加者の e ヘルスリテラシーのスコアとした。90 名の実験参加者の検索専門性のスコアは、平均 2.80, 標準偏差 0.75, 最小値 1.25, 最大値 4.50 であった。スコアの上位 20% の実験参加者 17 名を「e ヘルスリテラシーの高いユーザ」、残りの 80% にあたる 73 名を「e ヘルスリテラシーの高くないユーザ」とした。

また、検索専門性と e ヘルスリテラシーの関係についても調べた。90 名の実験協力者の検索専門性と e ヘルスリテラシーのスコアのスピアマンの相関係数を求めたところ 0.29 となり弱い正の相関が見られた。検索専門性、e ヘルスリテラシーそれぞれにおける 90 名の実験協力者の度数分布表は表 2 の通りであった。

4.2 ページのアノテーション

収集した検索クエリで検索されるページの適合性や質を分析するためのアノテーションを行った。本研究ではそれぞれの検索クエリで検索される検索結果の最上位 1 件を対象にアノテーションを行った。具体的には、Google が提供している Custom Search JSON API⁵を用いて各検索クエリの 1 位の検索結果を収集した。4 つのタスクで得られた 761 種類の検索クエリで最上位 1 件を収集し、280 件のページにアノテーションを行った。

収集したページに対して、以下のアノテーションを行った。

JAMA ベンチマーク基準: 評価基準は、Silberg らの JAMA ベンチマーク基準 [17] を使用した。JAMA ベンチマーク基準はウェブ上の健康情報の質を判断するための基準である。JAMA ベンチマーク基準として、以下の 4 つの基準がある。

- 著者情報: 情報の提供者、著者の所属している団体情報や保有している資格情報が記載してあるか。
- 参照: 参考文献や情報の出典元など参照した情報源が記載してあるか。
- 日付: 記事の更新日が記載されているか。
- 情報公開: ウェブサイトを運営していく上での資金を提供している企業が存在する場合に、その存在を明らかにしているか。

これら 4 つの基準それぞれについて、ページがその基準を満たしているか満たしていないかの 2 値でアノテーションを行った。

ページの適合性と意見: ページの内容が検索タスクに適合しているか、また、そのページに記載されている主張が、治療法が症状に効果があると主張しているかどうかという意見の観点でのアノテーションを行った。具体的には、(1) 検索タスクの効果を支持している、(2) 検索タスクの効果に反対している、(3) 1 と 2 両者の意見を含んでいる、(4) 検索タスクと関連しているが 1, 2, 3 どれにも当てはまらない、(5) 検索タスクと関係がないページである、という 5 種類のアノテーションを行った。この 5 種類のアノテーションを基に、1, 2, 3 に当てはまるものを、治療法と症状の関係を議論としているページとして検索タスクに対して適合であり、それ以外を不適合であると判断した。また、ページの意見の評価項目が 1 と 3 であるものを検索タスクとページの意見が一致していると判断し、それ以外を一致していないと判断した。

これらのアノテーションを著者らの所属する研究科の大学院生 3 名に行ってもらった。評価は 3 名独立して行い、3 名の多数決を取ることで JAMA ベンチマークの 4 基準、適合性、検索タスクとページの意見の一致のアノテーションとした。

4.3 トピック適合性とタスクとページの意見の一致

まず、RQ1 を明らかにするため、実験協力者から収集した検索クエリで検索される検索結果最上位 1 件が、検索タスクに関連したページであるかどうかを分析した。ここでは、トピック適合性と、タスクとページの意見の一致という 2 つの観点から分析を行った。トピック適合性は、表 2 に示した。例えば、タスク ID3 では『「てんかんに鍼治療が効果的ではない」と主張している発見したページのような、質の高いほかのページを探す』というタスクでは、てんかんと鍼治療の関係について議論しているページであれば適合であると考えた適合性である。

5: <https://developers.google.com/custom-search/v1/overview?hl=ja>

次に、タスクとページの意見の一致は、同タスクの場合、針治療がてんかんに対して効果がない、あるいはエビデンスが不足していると議論しているようなページのみ一致と考え、それ以外は一致していないと考える。トピック適合性がページの主張を考慮せず適合性を判断するのに対して、タスクとページの意見の一致はページの主張が検索タスクと一致している必要がある。トピック適合性とタスクとページの意見の一致はそれぞれ0または1の2値であり、4.2節で述べたアノテーションから求めた。

表3は各ユーザごとのトピック適合性とタスクとページの意見の一致をタスクごとにまとめた表である。今回の実験では、ページのアノテーション時にページが取得不可能な検索結果が含まれていたため、そのような検索クエリに関する値は欠損値として扱い、クエリレベルのマイクロ平均を用いた。まず、全ユーザのトピック適合率に着目すると、ID1が最も高く0.77であり、ID2が最も低い0.45となっていた。ID2が低いトピック適合率となった理由として考えられるのは、ユーザが2番目に多く考えた検索クエリである「不眠症 音楽」が適合するページにアクセスできていなかったからである。また、ID3はトピック適合率とタスクとページの意見の一致率の差が他のタスクよりも大きいことが分かった。これは、ウェブ上に検索タスクと一致する情報が少なかったことが理由として考えられる。

検索専門性やeヘルスリテラシーによる差異に着目すると、今回の実験では全ての検索タスクに共通して見受けられる傾向は得られなかった。たとえば、ID2では、検索専門性の高いユーザが「不眠症 音楽 研究 有効」という、「研究」というキーワードを含めた検索クエリを作成していたが、この検索クエリでは最上位1件の検索結果はタスクに適合していなかった。

4.4 ページの質

次に、ページがJAMAベンチマークの4基準をどの程度満たしているかを分析した。表4は各クエリで得られたトピック適合性を満たすページについて、JAMAベンチマークの各基準のマイクロ平均を求めたものである。表中の4基準の平均はこれら4基準の合計値を4で割った値である。

RQ1の健康情報検索における、ユーザが作成した検索クエリと得られるページの質はどのような関係か調べる。表4を見ると、情報公開に関する基準がどのタスクでも高いことが分かる。一方で、参照に関する基準は他の基準に比べてそれを満たすページの割合が少なく、情報の引用元を明記しているページが少ないことがわかる。以上からRQ1の答えとして、ユーザが得られるページには情報公開が含まれている可能性が最も高く、参照が含まれている可能性が最も低いという関係になった。

RQ2は検索専門性の有無によるページの質の変化に注目することであった。実際に検索専門性の高いユーザと低いユーザの得られるページの質を比較した場合に、全ての検索タスクに共通して見受けられる傾向は得られなかった。以上の結果から、RQ2の答えとして検索専門性の有無によって検索で得られるページの質は変化しなかった。

RQ3では健康情報に関するリテラシーの有無によるページ

の質の変化に着目をした。eヘルスリテラシーの高いユーザと低いユーザのページの質を比較した。結果として、健康情報に関するリテラシーの有無によって検索で得られるページの質は変化しなかった。個々の検索クエリを分析すると、例えば検索専門性の高いユーザがID1のタスクにおいて「高血圧 ココア 有効性 試験 結果」という検索クエリを入力しており、この検索クエリで検索されたページはJAMAベンチマーク基準のすべてを満たしており、質の高いページを得られる検索クエリであったと言える。ID2においては検索専門性の高いユーザが「不眠症 音楽 効果 立証 成人」という検索クエリを入力しており、これもJAMAベンチマーク基準のすべてを満たしていた。一方で、こうした専門性の有無によりそのような検索クエリの傾向の顕著な違いは見受けられなかった。

4.5 検索クエリの表層的な分析

最後に、各ユーザ属性ごとに入力された検索クエリの違いを見るため、今回の実験で得られた検索クエリの表層的な分析を行った。具体的には、実験で入力された検索クエリの単語数を分析した。検索クエリの単語数とは、空白で区切られた単語の数を指す。たとえば、「高血圧 ココア 効果」という検索クエリであれば単語数は3となる。一般的に検索専門性の高いユーザはそうでないユーザよりも短い検索クエリを投入し目的を達成できることが知られている[19]。また、健康情報の質の判断に対する知識があるユーザは、専門的なキーワードを検索クエリに含めることが出来るかもしれない。このように、ユーザ属性により検索クエリに違いが出てくるのではないかと考えた。

表5はユーザ属性ごとに検索クエリの単語数をタスクごとにまとめた表である。今回の実験ではどのタスクも2.70から2.80前後となっていた。これは、タスクID1であれば「高血圧 ココア 効果」のように症状と治療法の効果の有無を問うような検索クエリが多かったためであると考えられる。

次に、検索専門性の高いユーザと検索専門性の低いユーザの単語数を比較する。検索専門性の高いユーザの単語数としては、4つのタスクで2.65から2.80となった。検索専門性の低いユーザの場合では、4つのタスクで2.71から2.86となり、大きな差は見受けられなかった。同様に、eヘルスリテラシーについてもその差異による単語数の違いは見受けられなかった。

5 議論

本研究での結果についての考察について述べる。

5.1 考察

本研究では検索専門性やeヘルスリテラシーにより得られるページの質が変わると考え分析を行った。一方、表3や表4にあるように、今回の実験からはこれらの属性の違いによる影響は見受けられなかった。その理由の1つとして、検索クエリが具体的であるためにトピック適合性やタスクと意見との一致を満たさないページが得られたということがあげられる。たとえば、ID1のタスクにおいて「フラバノール高含有 ココア 有意

表 3 トピック適合率とタスクとページの意見の一致率. () 中の数値は標準偏差を表す. なお, 表中のタスク ID は表 1 で示した ID に対応している.

タスク ID	ユーザの属性	トピック適合率	タスクとページの意見の一致率
ID1	全ユーザ	0.77 (0.42)	0.77 (0.42)
	検索専門性が高いユーザ	0.70 (0.46)	0.70 (0.46)
	検索専門性が高いユーザ	0.79 (0.41)	0.79 (0.41)
	e ヘルスリテラシーが高いユーザ	0.65 (0.48)	0.65 (0.48)
	e ヘルスリテラシーが高いユーザ	0.80 (0.40)	0.80 (0.40)
ID2	全ユーザ	0.45 (0.50)	0.45 (0.50)
	検索専門性が高いユーザ	0.37 (0.49)	0.37 (0.49)
	検索専門性が高いユーザ	0.47 (0.50)	0.47 (0.50)
	e ヘルスリテラシーが高いユーザ	0.49 (0.39)	0.39 (0.49)
	e ヘルスリテラシーが高いユーザ	0.47 (0.50)	0.46 (0.50)
ID3	全ユーザ	0.61 (0.49)	0.12 (0.32)
	検索専門性が高いユーザ	0.66 (0.48)	0.14 (0.35)
	検索専門性が高いユーザ	0.60 (0.49)	0.11 (0.32)
	e ヘルスリテラシーが高いユーザ	0.55 (0.50)	0.16 (0.37)
	e ヘルスリテラシーが高いユーザ	0.62 (0.49)	0.11 (0.31)
ID4	全ユーザ	0.71 (0.45)	0.56 (0.50)
	検索専門性が高いユーザ	0.65 (0.48)	0.51 (0.50)
	検索専門性が高いユーザ	0.73 (0.45)	0.57 (0.50)
	e ヘルスリテラシーが高いユーザ	0.75 (0.44)	0.55 (0.50)
	e ヘルスリテラシーが高いユーザ	0.71 (0.46)	0.57 (0.50)

表 4 トピック適合性を満たすページに対する JAMA ベンチマーク基準のマイクロ平均. () 中の数値は標準偏差を表す. なお, 表中のタスク ID は表 1 で示した ID に対応している.

タスク ID	ユーザ属性	著者情報	参照	日付	情報公開	4 基準の平均
ID1	全ユーザ	0.57 (0.50)	0.47 (0.50)	0.83 (0.38)	0.98 (0.15)	0.71 (0.28)
	検索専門性が高いユーザ	0.54 (0.51)	0.54 (0.51)	0.83 (0.38)	0.97 (0.17)	0.72 (0.28)
	検索専門性が高いユーザ	0.57 (0.50)	0.45 (0.50)	0.83 (0.38)	0.98 (0.15)	0.71 (0.28)
	e ヘルスリテラシーが高いユーザ	0.52 (0.51)	0.42 (0.50)	0.91 (0.29)	0.97 (0.17)	0.70 (0.27)
	e ヘルスリテラシーが高いユーザ	0.57 (0.50)	0.48 (0.50)	0.81 (0.39)	0.98 (0.15)	0.71 (0.28)
ID2	全ユーザ	0.74 (0.44)	0.60 (0.49)	0.71 (0.46)	0.99 (0.09)	0.76 (0.29)
	検索専門性が高いユーザ	0.63 (0.50)	0.58 (0.51)	0.63 (0.50)	1.00 (0.00)	0.71 (0.34)
	検索専門性が高いユーザ	0.76 (0.43)	0.60 (0.49)	0.72 (0.45)	0.99 (0.10)	0.77 (0.28)
	e ヘルスリテラシーが高いユーザ	0.90 (0.31)	0.65 (0.49)	0.70 (0.47)	1.00 (0.00)	0.81 (0.24)
	e ヘルスリテラシーが高いユーザ	0.71 (0.46)	0.59 (0.49)	0.71 (0.46)	0.99 (0.10)	0.75 (0.30)
ID3	全ユーザ	0.23 (0.42)	0.13 (0.34)	0.25 (0.44)	1.00 (0.00)	0.40 (0.27)
	検索専門性が高いユーザ	0.21 (0.42)	0.15 (0.36)	0.27 (0.45)	1.00 (0.00)	0.41 (0.29)
	検索専門性が高いユーザ	0.23 (0.42)	0.13 (0.34)	0.25 (0.43)	1.00 (0.00)	0.40 (0.27)
	e ヘルスリテラシーが高いユーザ	0.32 (0.48)	0.21 (0.42)	0.32 (0.48)	1.00 (0.00)	0.46 (0.32)
	e ヘルスリテラシーが高いユーザ	0.21 (0.41)	0.12 (0.32)	0.24 (0.43)	1.00 (0.00)	0.39 (0.26)
ID4	全ユーザ	0.65 (0.48)	0.81 (0.39)	0.88 (0.33)	0.97 (0.18)	0.83 (0.24)
	検索専門性が高いユーザ	0.67 (0.48)	0.88 (0.33)	0.91 (0.29)	1.00 (0.00)	0.86 (0.20)
	検索専門性が高いユーザ	0.64 (0.48)	0.80 (0.40)	0.87 (0.34)	0.96 (0.19)	0.82 (0.25)
	e ヘルスリテラシーが高いユーザ	0.66 (0.48)	0.71 (0.46)	0.95 (0.23)	1.00 (0.00)	0.83 (0.22)
	e ヘルスリテラシーが高いユーザ	0.64 (0.48)	0.84 (0.37)	0.86 (0.35)	0.96 (0.20)	0.83 (0.25)

差」という検索クエリが e ヘルスリテラシーが高いユーザによって入力されていたが, この検索クエリではタスクと適合するページが得られなかった.

今後, ユーザ属性の違いにより入力される検索クエリにどのような違いがあるのか分析を進める必要がある.

5.2 実験の限界点

本研究での実験における限界点を述べる.

5.2.1 タスク文から検索クエリへの影響

タスク文に含まれる単語から検索クエリへ影響している可能性がある. 本研究では, ユーザの考えられる検索クエリを調査する際に, ある症状に対して治療法が有効かどうかというタスク文をユーザに読んでもらい検索クエリを作成している. その

表 5 ユーザ属性ごとの検索クエリ中の単語数. () 中の数値は標準偏差を表す. なお, 表中のタスク ID は表 1 で示した ID に対応している.

ユーザ属性	タスク ID			
	ID1	ID2	ID3	ID4
全ユーザ	2.81 (0.83)	2.70 (0.81)	2.86 (0.72)	2.81 (0.72)
検索専門性が高いユーザ	2.78 (0.92)	2.65 (0.87)	2.80 (0.85)	2.75 (0.83)
検索専門性が低いユーザ	2.82 (0.81)	2.71 (0.80)	2.86 (0.69)	2.82 (0.70)
e ヘルスリテラシーが高いユーザ	2.64 (1.01)	2.40 (0.92)	2.60 (0.88)	2.55 (0.88)
e ヘルスリテラシーが低いユーザ	2.85 (0.78)	2.77 (0.77)	2.90 (0.67)	2.87 (0.67)

ため, ユーザが作成する検索クエリがタスク文の内容に大きな影響を受けている. たとえば, 「ココア 高血圧 効果」という検索クエリはその全てがタスク文に含まれている.

今後は, ユーザが入力した検索クエリをより詳細に分析し, タスク文に含まれていないキーワードがどの程度含まれているか, また, 例示したコクランのページに含まれているキーワードがどの程度あるか, といった分析を進めるとともに, これらの影響を生じさせないような検索タスクの設計についても考える必要がある.

6 ま と め

本研究では, ユーザがウェブ検索時に作成する検索クエリに着目し, 検索クエリがどの程度効果的に質の高い健康情報を検索可能なかを調査した. 90 名の実験協力者に 4 つの検索タスクを与え検索クエリを入力してもらった実験を行うことで検索クエリを収集し分析を行った.

分析の結果, 各検索クエリ最上位 1 件のページのトピック適合率は 0.45 から 0.77 であり, 検索タスクとページの意見が一致するかどうかは 0.12 から 0.77 と個々の検索タスクによって大きな差があった. また, JAMA ベンチマーク基準を用いてページの質を分析した結果, 多くのページは情報公開を満たしていた一方で, 参照を満たすページは他の基準に比べて少ないことが分かった.

また, 検索専門性や e ヘルスリテラシーの有無による検索クエリの違いは今回の実験からは見受けられなかった. 今後は分析を進め, ユーザの違いによって作成される検索クエリが具体的にどのように異なるのかを分析する予定である.

謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP18H03494, JP18H03243, ならびに, 木下基礎科学研究基金助成事業による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] A. Aula, R. M. Khan, and Z. Guan. How does search behavior change as search becomes more difficult? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 35–44, 2010.
- [2] L. Baker, T. H. Wagner, S. Singer, and M. K. Bundorf. Use of the internet and e-mail for health care information: results from a national survey. *JAMA*, 289(18):2400–2406,

- 2003.
- [3] F. Beck, J.-B. Richard, V. Thanh-Nguyen, I. Montagni, I. Parizot, and E. Renahy. Use of the internet as a health information resource among french young adults: results from a nationally representative survey. *Journal of Medical Internet Research*, 16(5):e128, 2015.
- [4] C. S. Bond. Web users’ information retrieval methods and skills. *Online Information Review*, 28:254–259, 2004.
- [5] D. Charnock, S. Shepperd, G. Needham, and R. Gann. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology & Community Health*, 53(2):105–111, 1999.
- [6] A. Cipriani, T. Furukawa, and C. Barbui. What is a cochrane review? *Epidemiology and Psychiatric Sciences*, 20(3):231–233, 2011.
- [7] Y. Goto, I. Sekine, H. Sekiguchi, K. Yamada, H. Nokihara, N. Yamamoto, H. Kunitoh, Y. Ohe, and T. Tamura. Differences in the quality of information on the internet about lung cancer between the united states and japan. *Journal of Thoracic Oncology*, 4(7):829–833, 2009.
- [8] B. W. Hesse, D. E. Nelson, G. L. Kreps, R. T. Croyle, N. K. Arora, B. K. Rimer, and K. Viswanath. Trust and sources of health information: the impact of the internet and its implications for health care providers: findings from the first health information national trends survey. *Archives of Internal Medicine*, 165(22):2618–2624, 2005.
- [9] C. Liu, X. Zhang, and W. Huang. The exploration of objective task difficulty and domain knowledge effects on users’ query formulation. 53(1):1–9, 2016.
- [10] M. Liu, Y. Liu, J. Mao, C. Luo, M. Zhang, and S. Ma. “Satisfaction with failure” or “unsatisfied success” investigating the relationship between search success and user satisfaction. In *Proceedings of the 2018 World Wide Web Conference*, pp. 1533–1542, 2018.
- [11] M. McMullan. Patients using the internet to obtain health information: how this affects the patient–health professional relationship. *Patient Education and Counseling*, 63(1-2):24–28, 2006.
- [12] K. Nakayama, W. Osaka, T. Togari, H. Ishikawa, Y. Yonekura, A. Sekido, and M. Matsumoto. Comprehensive health literacy in japan is lower than in europe: a validated japanese-language assessment of health literacy. *BMC Public Health*, 15(1):505, 2015.
- [13] K. Nemoto, H. Tachikawa, N. Sodeyama, G. Endo, K. Hashimoto, K. Mizukami, and T. Asada. Quality of internet information referring to mental health and mental disorders in japan. *Psychiatry and Clinical Neurosciences*, 61(3):243–248, 2007.
- [14] C. D. Norman and H. A. Skinner. eHEALS: the eHealth literacy scale. *Journal of Medical Internet Research*, 8(4):e27, 2006.
- [15] S. Pothirattanachaikul, T. Yamamoto, Y. Yamamoto, and M. 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, pp. 1653–1662, 2019.

- [16] S. Pothirattanachaikul, T. Yamamoto, Y. Yamamoto, and M. Yoshikawa. Analyzing the effects of “people also ask” on search behaviors and beliefs. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pp. 101–110, 2020.
- [17] W. M. Silberg, G. D. Lundberg, and R. A. Musacchio. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor—let the reader and viewer beware. *JAMA*, 277(15):1244–1245, 1997.
- [18] E. Sillence, P. Briggs, L. Fishwick, and P. Harris. Trust and mistrust of online health sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 663–670, 2004.
- [19] R. W. White and D. Morris. Investigating the querying and browsing behavior of advanced search engine users. In *SIGIR*, pp. 255–262, 2007.
- [20] 光武誠吾, 柴田愛, 石井香織, 岡崎勘造, 岡浩一郎. eHealth literacy scale(eHEALS) 日本語版の開発. *日本公衛誌*, 58(5):361–371, 2011.
- [21] 厚生労働省. 平成 26 年版厚生労働白書 健康長寿社会の実現に向けて ～健康・予防元年～第 1 部第 2 章「健康をめぐる状況と意識」. 平成 26 年度厚生労働白書, 2014.
- [22] 山本祐輔, 山本岳洋, 大島裕明, 川上浩司. ウェブアクセスリテラシー尺度の開発. *情報処理学会論文誌 データベース*, 12(1):24–37, 2019.
- [23] 石丸知宏, 宮内健悟, 桑原恵介, 服部理裕, 根本博. インターネットのうつ病の診断に関する情報は信頼できるか? *労働科学*, 94(1):12–18, 2018.