第21回年次大会予稿

ユーザ種別と閲覧時間を考慮したWeb情報検索の行動特性調査

Experiment on Action Characteristics of Web Information

Retrieval Based on User Types and Browsing Time

遠藤淳一^{1*}, 村川猛彦²

Jun'ichi ENDOU^{1*}, Takehiko MURAKAWA²

1 和歌山大学大学院システム工学研究科

Graduate School of Systems Engineering, Wakayama University

〒640-8510 和歌山市栄谷930

2 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

〒640-8510 和歌山市栄谷930

*連絡先著者 Corresponding Author

Web検索の履歴や検索中のユーザの行動を分析することで、ユーザの検索目的や行動特性を把握し、Web情報推薦に役立てようとする研究が盛んである。これまでの研究でも検索活動にかかった時間やWebページ閲覧時間について調査したものは存在するが、それらに影響を与えられるユーザの知識経験やページの質との関連を調べたものは見られない。そこで本研究ではユーザの特性やページの質と、Webページ閲覧時間の関係に焦点を当てて個人のWeb情報検索の過程を取得・分析した。本稿では実施した実験及びその結果について述べる。

The researches that analyze the search histories of Web pages associated with the users' actions become popular to grasp the purpose of the search or to clarify the characteristics of the behavior. Although some works thought of the time of searching and browsing, those did not take into consideration the users' experiences, the knowledge or the page qualities. The goal of our research is to obtain the process of personal Web searching, focusing on relations between the users' characteristics or the page qualities and the times of information retrieval. This paper presents the overview and the results of the experiments.

キーワード: 情報推薦, インターネット, 閲覧時間, 閲覧履歴

Information Retrieval, Internet, Browsing Time, Browsing History

1. まえがき

インターネット上には大量の情報が存在しており、ユーザは Yahoo! や Google などの検索エンジンを用いて情報を検索する。ユーザが訪れたページは閲覧履歴として保存されるが、この閲覧履歴を分析し、ユーザの検索目的や興味を把握し、情報推薦に利用しようとする研究が盛んである。

高久ら[1]はユーザの情報検索中の眼球運動や閲覧履歴などの様々なデータを分析し、ユーザが情報をどのようにピックアップしどのような行動が選択されるか、その際どこを注視するかなどを調査し、ユーザごとの行動の傾向を示した。

閲覧履歴を分析する際に主に利用されるのは「検索キーワード」「Web ページ閲覧時間」「移動先のページ URL」などであるが、特に Web ページ閲覧時間について考慮している研究に清水[2]や三原ら[3]の研究がある。しかし Web ページ閲覧時間はユーザの興味関心の他にも、ページの良し悪しや情報量などに影響される。また検索活動全体にかかる時間はユーザの情報検索経験や調査課題に対する前知識の有無などに影響を受ける。しかしながら、既存研究ではそれらのユーザおよびページの情報がどのように閲覧時間に影響を与えるかはあまり考慮されていない。それらを考慮して検索過程を調査することで、より優れた Web ページ推薦システムを構築する指針の導出が期待される。

本研究では、ユーザの特性やページの質などの様々な要因と Web ページ閲覧時間の関係を調査することを目的として個人の Web 情報検索の過程を取得する実験を行い、結果を分析した。

2. 実験

2.1 実験の目的および概要

Herlocker ら[4]によると、検索推薦の目的は 4 つに大別され、ほとんどの推薦システムはそのうちの 2 つ、具体的には

①適合するアイテム（検索活動においては Web ページ）の発見

②何らかのフォーマットに従いアイテムが整列されている際の、評価の閲覧のどちらかをユーザの検索目的であると想定している。

そこで、今回はこの 2 種を実験の課題設定の対象とした。すなわち、①に相当する課題としてエルトゥール号遭難事件に対する課題（以下「事件課題」）を、②に相当する課題として行ってみたい旅行先についての課題（以下「旅行課題」）の 2 つの課題を設定し実験を行った。また検索活動の経験差、調査課題に対する前知識の有無、ユーザの興味などにより、閲覧時間に差が生じると考えられるため、情報検索に慣れているユーザとして和歌山大学大学院システム工学研究科に所属する大学院生 4 名に、あまり情報検索に慣れていないユーザとして和歌山大学経済学部・観光学部の大学生 5 名に協力を依頼した。本稿では紙面の都合上、主に事件課題について詳しく述べ、旅行課題については結果のみを簡潔に記述する。

2.2 課題について

事件課題は大学での小レポートを想定し、1890 年に和歌山県串本町沖で発生したエルトゥール号遭難事件について、簡単な概要と、事件がその後の国交にどのような影響を与えたのかを調査するものである。各参加者はレポートを書く際に有用である

と思われる Web ページを検索し、概要を把握してもらう。ただしレポートの作成そのものは行わない。調査終了後にインタビューを実施し、実験前から事件について知っていたかどうかや有用と思ったページはどれかを尋ねた。この事件を題材としたのは、テレビ番組や和歌山県内の講演での題材として紹介されることがあり、実験協力者の専攻を問わず既知・未知の差が発生すると考えられたためである。

2.3 実験環境

実験では筆者らの研究室のデスクトップ PC を使用した。実験記録のための画面キャプチャの録画には「アマレココ」を、閲覧履歴の取得には Web ブラウザの Firefox のアドオンである「FromWhereToWhere」を使用した。そのため、実験協力者には Web ブラウザは Firefox のみを使うよう指示した。また、普段と同じような検索がしやすいように Yahoo! JAPAN と Google をあらかじめブックマークしておいた。実験後、参加者に今回使用した PC の環境は普段使用している検索の環境（Web ブラウザや検索エンジン、ブックマークなど）と比べてどのように感じたかを質問したが、全員が問題なく検索が行えたと回答した。

2.4 実験内容

実験は筆者らの研究室内で 1 人ずつ実施した。まず実験全体の流れを説明し、その後各参加者は事件課題について Web ブラウザを用いて調査を行わせた。課題には時間制限を設けず、ユーザが十分だと判断した時点で終了とした。事件課題が終了した後に事件についての前知識量や検索経験について質問するとともに、画面キャプチャの

録画を再生しながらインタビューを実施した。その際に有用だと感じたページ（欲しい情報が多かったページ）を聞き取った。これは 1 つでも複数でも良いものとし、なぜそう判断したかや検索行動中の疑問点についても尋ねた。その後旅行課題についても同様の流れで実験を実施し、両方の課題が終了した後に PC やインターネットの使用経験、実験で使った検索環境について質問した。実験終了後に画面キャプチャの録画と閲覧履歴のデータをもとに、各ページの閲覧時間をまとめた。

2.5 閲覧時間の取得

閲覧時間を求めるため、閲覧開始時刻と閲覧終了時刻を取得する必要がある。本稿では、それぞれ以下のようにした。ブラウザ上で Web ページが完全に表示され内容を読むことができるようになった時点を開覧開始時刻とする。ブラウザの「戻る」ボタンを押す、ページ内のリンクから別のページにアクセスするなどにより、現在の Web ページから移動しようとした時点を開覧終了時刻とする。タブ機能や別のウィンドウで Firefox を起動した場合は、閲覧するタブやウィンドウをアクティブにした時点を開覧開始時刻とし、他のタブやウィンドウをアクティブにした時点を開覧終了時刻とする。

ページの閲覧時間は URL のドメイン名でまとめ、同一のページを複数回閲覧した場合は、それぞれの閲覧時間を加算する。また検索語の入力時間や旅行課題における別紙への手書きの時間などは別途計測し、閲覧時間には含めないようにした。

3. 実験結果と考察

3.1 ページの質と閲覧時間の関係

結果を述べる前に、用語を定義する。Web 情報検索中の検索結果が表示されるページ (Search Engine Result Page) を SERP と記す。SERP 以外のページを non-SERP と記す。

この節では、有用だったとユーザに判断されたページを質の良いと思われたページとし、実験協力者ごとに良いと思われたページの閲覧時間と、non-SERP の閲覧時間の平均を比較した結果について記す。良いと思われたページは、全ページの閲覧時間平均と比較するとどれだけ長く閲覧されているかを調査した。

事件課題における結果を表 1 に示す。A~I は実験協力者である。良いと思われたページの閲覧時間は、全ページの閲覧時間平均よりも平均して 2.2 倍ほど長く閲覧されていた。そのため「質の良いページは他のページに比べ閲覧時間が長くなる」という結論が導出できた。

旅行課題でも同様の傾向が見られたが、全員に当てはまる傾向ではなかった。また旅行課題で良いと思われたページの閲覧時間は事件課題に比べ閲覧全体に占める割合が高かった。これは事件課題に比べてページの良し悪しを判断するのに時間がかかり、かつ良いと思ったページですべての調査を

済ませようとする傾向があったためである。

また事件課題において、9 名中 7 名の被験者に「最初に閲覧したページの閲覧時間が長くなる」という傾向が確認された。更にその 7 名中 5 名は、最初に閲覧したページを良いページであると回答した。加えて、最初に閲覧されるページは検索結果の上位に存在することが多かった。これは事前知識の有無を問わず、多くの被験者にまず適当なページを閲覧し事件の概要を把握しようとする傾向があったためである。そのページで十分に情報が得られることも多かったため、このような結果になったと推測できる。以上の傾向をまとめた Web 情報検索行動のモデルを図 1 に示す。実際の行動としては、検索語 1 に「エルトゥールル号遭難事件」、検索語 2 に「エルトゥールル号 トルコ」など、検索語 1 と似た言葉や、検索語 1 で訪れたページに記載されていた言葉が使用されることが多かった。

3.2 ユーザ特性と閲覧時間の関係

この節では「入力された検索語の総数」「SERP, non-SERP 閲覧数」「それぞれの閲覧時間合計」「それぞれの閲覧時間平均」の 5 つについて、専攻などのユーザの特性で参加者を 2 つのグループに分けた際にどのようなことが言えるかを確認した。

表 1 ページの質と閲覧時間について(事件課題)

	A	B	C	D	E	F	G	H	I
①良いページの 閲覧時間(秒)	81.00	120.00	115.00	50.00	159.00	64.00	119.00	487.00	22.00
②non-SERP 閲 覧時間平均(秒)	40.43	63.50	74.50	36.00	111.67	19.45	101.20	88.88	14.30
① ÷ ②	2.00	1.89	1.54	1.39	1.42	3.29	1.18	5.48	1.54

図1 検索行動のモデル

3.2.1 専攻による分類

実験協力者を経済・観光学部の大学生 5 名と、システム工学研究科大学院生 4 名に分けた場合の各項目の平均を表 2 に示す。

システム工学を専攻しているグループのほうが少ない検索語で多くのページを閲覧していた。また、どちらのグループも SERP よりも通常の Web ページを長時間閲覧していたが、システム工学を専攻する学生のほうがその傾向がより顕著に見られた。これは彼らがタブ機能などを活用し、あらかじめ興味のあるページを複数開いた後にそれらを順番に閲覧することが多く、検索結果表示ページをあまり閲覧しなかったからであると考えられる。

旅行課題においても、検索語総数以外の項目について同様の傾向が見られた。

3.2.2 事件の前知識の有無による分類

実験参加者を、アンケートにおいて本事件についての前知識が全くなかったか名前程度しか知らないと回答した 4 名と、実験前から概要を知っていたと回答した 5 名に分けた場合の各項目の平均を表 3 に示す。

検索語の総数平均、検索結果表示ページの閲覧数以外は専攻で分類した際と同様の傾向が見られた。前知識のある参加者の検索語総数が多いのは、より多くの関連する語について検索を行っていたためである。しかしながら、検索にかかった総時間はあまり大きな差がなく、前知識のある協力者

表 2 専攻による分類(事件課題)

	経済・観光	システム工学
検索語の総数(個)	3.40	2.25
SERP 閲覧数(ページ)	7.80	5.50
non-SERP 閲覧数(ページ)	4.60	8.50
SERP 閲覧時間合計(秒)	144.20	79.50
non-SERP 閲覧時間合計(秒)	265.60	393.50
SERP 閲覧時間平均(秒)	18.49	14.45
non-SERP 閲覧時間平均(秒)	57.74	46.29

表 3 前知識の有無による分類(事件課題)

	なし	あり
検索語の総数(個)	2.40	3.50
SERP 閲覧数(ページ)	6.20	7.50
non-SERP 閲覧数(ページ)	4.20	9.00
SERP 閲覧時間合計(秒)	132.60	94.00
non-SERP 閲覧時間合計(秒)	310.20	337.75
SERP 閲覧時間平均(秒)	21.39	12.53
non-SERP 閲覧時間平均(秒)	73.86	37.53

のほうが検索結果表示ページおよび通常の Web ページの両方の閲覧時間平均が短くなった。以上より、前知識のあるユーザは短

時間で効率よく検索を行っていることが確認できた。

旅行課題においては、前知識の代わりに旅行先調査の経験の有無によって分類したが、明確な差は見られなかった。これは専攻や経験以上に検索の難易度の影響を強く受けたためであると思われる。旅行課題での調査目的地は協力者自身が定めたが、Web上に情報が多く存在する目的地は短時間で情報が集められ、情報量が少ない目的地は検索に時間がかかることが多かった。

3.3 考察

以上の結果より、ユーザに良いと思われるページの閲覧時間は、他のページに比べて長くなりやすいことが確認できた。一方で、良いと思われるページはユーザが最初に閲覧するページであることが多く、それは検索結果の最上位に存在するページであることも多かった。よって、他のユーザが長時間閲覧したページの推薦と通常のWeb検索は、それぞれの結果の最上位が同じような結果になることが考えられる。また、ユーザの閲覧時間は検索課題に対する前知識量や難易度、そして個人の文書の閲覧速度にも影響を受けるため、システム化の際は考慮が必要である。

4. あとがき

本稿ではユーザの特性とWebページ閲覧時間に着目した情報検索に関して実施した実験について報告した。ユーザに良いと思われるページの閲覧時間は長くなること、最初に閲覧されるページの閲覧時間は長くなること、閲覧時間は個人の文書を読む速度や検索目的・難易度に大きな影響を受け

るため、システム化の際は考慮が必要であることが確認された。

筆者らの研究室では、Webページ及び検索語の推薦システムを試作してきた [5]。今後は本稿で示した結果を取り入れながら実用的な情報推薦システムを構築・運用し、評価を行っていく予定である。

参考文献

- [1] 高久雅生；江草由佳；寺井仁；齋藤ひとみ；三輪眞木子；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，Vol.20，No.3，pp.249-276，2010.
- [2] 清水勇喜：「WWWにおけるユーザの興味対象と閲覧時間の関係の調査」，情報処理学会第57回全国大会，No.3，pp.191-192，1998.
- [3] 三原宏一郎；寺邊正大；橋本和夫：「ページ閲覧時間を考慮したWebログマイニング手法の提案」，情報処理学会研究報告 知能と複雑系，Vol.2007，No.67，pp.39-44，2007.
- [4] Herlocker, Jonathan L；Konstan, Joseph A；Terveen, Loren G；Riedl, John T："Evaluating Collaborative Filtering Recommender Systems"，ACM Transactions on Information Systems，Vol.22，No.1，pp.5-53，2004.
- [5] 廣田勇希：「閲覧時間を考慮したWeb情報推薦システムの構築」，和歌山大学システム工学部2012年度卒業論文，22p，2013.