

Web 閲覧中のユーザの視線に基づく関連情報提示システムの開発及び評価

Development and Evaluation of a System for Presenting Related Information Based on the User's Gaze while Browsing the Web page

森 大河[†]星野 祐子[†]石井 英里子[‡]山田 光穂[‡]

Taiga Mori

Yuko Hoshino

Eriko Ishii

Mitsuho Yamada

1. はじめに

2014 年, Tobii Tech と EyeTribе によって, 最初の安価な非接触型の視線入力機がリリースされ, アイトラッキング技術の注目が高まってきている. 古くから視線情報からの興味や注意を推測できると言われており, これらアイトラッキング技術による注視位置の解析は Search Engine Optimization(SEO) 対策にも用いられている. また, 以前であれば旅行を計画する際は観光ガイドブックなどにより, 旅行先近辺の施設についての情報を得ることが一般的であった. 情報通信技術の普及によってインターネットが「趣味・娯楽に関する情報を得る」ために最も利用するメディアとして, 60 代以上を除く各年代で書籍, 雑誌, テレビによる情報収集を上回り[1], 情報収集の多くが web 閲覧によるものへと変化していった. こうした背景を踏まえ, 今回我々は web 閲覧中のユーザーの視線を活用し, そのページに関連する情報を検索クエリとして提示する関連情報提示インタラクションの開発を試みた. これにより, web 閲覧による観光, 旅行に関する情報収集の支援や, ユーザーが知らなかった意外な施設の提示などが可能となり, 様々な関連情報を容易に取得することができることから, ユーザーの web ブラウジングをサポートできると考える. また利用者によるアンケート調査から本システムの精度評価を行い, 複数の web ページ間で本システムによるインタラクションの精度に差が生じるかについても検証を行った.

2. 関連研究

閲覧 web ページからキーワードを抽出し情報収集を支援しようとする試みは以前から存在する. 渡辺らの研究[2]では本文中に存在する固有表現を抽出しキーワード候補とする. それらに対して Web 上でどの程度頻出する単語かを計算したスコアである WebIDF[3]によるキーワードスコアリングを施し, スコア上位のキーワードをユーザーに通知し提示する検索支援 UI の開発を行っている.

上村らの研究[4]では閲覧 web ページより抽出したキーワードを利用することで, 関連サイトの表示や, 検索クエリの入力候補の提示を行い, web 閲覧を補助するインターフェースを実現する手法を示している.

このように閲覧 web ページからキーワードを抽出し情報収集を支援しようとする試みは以前から存在するが, 本研究ではさらに web 閲覧者の視線情報を利用することで, よりパーソナライズ化された関連情報の提示が行えると考え, 開発を試みた.

3. 視線入力デバイス

本章では本研究で使用した安価なコンシューマー向け非接触型視線入力装置である Tobii Eye Tracker 4C について述べる. 当機材は眼球運動計測手法として角膜反射法(PCCR)が用いられている. 角膜反射法はまず角膜上に光の

反射点を生じさせて, その画像をカメラで撮影し, 撮影された眼球の画像から角膜上の光の反射点と瞳孔を識別する. 光の反射点やその他の幾何学的特徴を基に眼球の方向が算出される. サンプルレートは 90Hz[5]であり, 11.1ms ごとに視線の抽出が可能となっている.



図 1. Tobii Eye Tracker 4C

4. インタラクションシステム

本システムは, まずユーザーの視線に基づき, 閲覧中の web ページのテキスト要素を自動で取得し, 文章解析や単語解析を行う. その後, 導き出されたテキストのコア情報を再構成して検索クエリ化し, ユーザーに提示を行うことで閲覧した web ページに関する関連情報の検索を行うことができる. 本システムは 3 つのフェーズに分類することができ, まずユーザーの注視点に基づき閲覧文章を取得する文章取得フェーズ, 次に tf-idf という重要語の評価指標を用いた文章解析を行う文章解析フェーズ, 最後に単語のベクトル表現が可能な word2vec を用いた単語解析を行う内容把握フェーズという 3 つのフェーズによって導き出されたコア情報が検索クエリとしてユーザーに提示される.

4.1 文章取得フェーズ

本フェーズではテキストを取得して次のフェーズで文章の解析を行うまでの処理を行っている. Selenium を経由して Google Chrome(以下, Chrome と表記)を開く必要があるため, ベースとなる部分を C# と Windows Presentation Foundation(WPF)を用いて GUI アプリケーションとして作成した. 図 2 にこの GUI アプリケーションの画面を示す. この画面からボタンをクリックするのみでスク립トが埋め込まれた Chrome を開くことができる. システムは視線に基づいて JavaScript を動作させて, 閲覧ページから一文単位でテキストを取得する. なお, 視線座標から文章の取得を行うために未読の文章は取得を行っていない. またコンシューマー向けの視線入力機でユーザーが注視している一文字を完全に特定する場合, 装置の精度が十分ではなくフォントサイズがユーザーの読み進めに大きな影響を及ぼすこと等を考慮した結果から, 一文字単位ではなくより広い範囲から対象の HTML 要素を取ることができ一文単

位での取得を選択している。一文を取得後は一度、文章解析処理にデータを送信して形態素解析を行い、文章の内容を構成する要素だけに変換する。形態素解析によって分かち書き形式に変換されたテキストデータから抽出された必要情報は単語単位で出力されるため、システムはこれを一時的に保管する。ページ閲覧が終わるまで入力される文章の解析を行い、閲覧終了後にまとめて文章解析フェーズの処理を行う。また、著者らによる先行研究[6]では文章の読み度合いを考慮しておらず、文章に視線が向いているか向いていないかのみで評価を行っていた。そこで本研究では熟読中に急激な注視点の変化は生じないという仮定のもと、日本語文章の平均の読み速度と web ページの推奨 px 数から算出した閾値により、流し読みと判定された文章は抽出されないように改善を加えた。日本語文章の平均の読み速度については、168 名の大学生を対象に文章の読み速度の計測を行った齋田による研究[7]では平均速度 504.9 文字/分、31 名の大学生を対象に読み速度の計測を行った小林らの研究[8]では平均 737 文字/分と報告されている。今回はこれらの研究を参考に両者の中間である平均 600 文字/分とした。また、web ページの推奨 px 数は株式会社アイ・エム・ジェイの調査[9]によるタブレットユーザーが読みやすいと感じるフォントサイズ 16px とした。今回はこれらの値をもとに $10 \text{ 文字/s} \times 16\text{px} = 160\text{px}$ という閾値を設定し、システムにより 1 秒ごとに取得した注視点の座標が、この閾値以上に視線を動かしていた場合は文章を読んでいない、流し読みを行った(平均の読み速度より速い)と判定し、文章抽出を行わないように設定した。

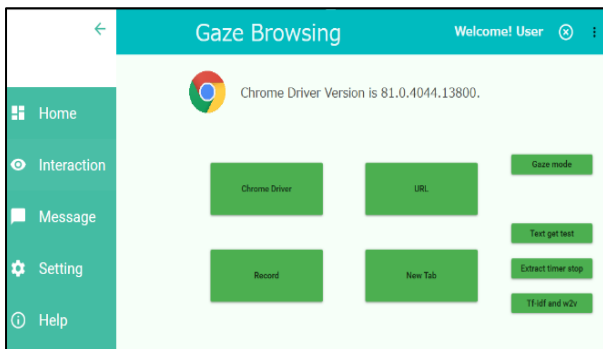


図 2 作成した GUI アプリケーション

4.2 文章解析フェーズ

文章解析の処理は先ほどの文章取得フェーズと並行して閲覧中に行われる処理とページ閲覧の終了後に行われるものの 2 種類が存在する。閲覧中に行われている処理では、入力された一文に対して MeCab を用いた形態素解析を行い、文章の内容を実質的に構成するとされる内容語である名詞、動詞、形容詞、副詞から文章についてコアとなる情報を持たない副詞以外の単語を分かち書きで取得する。なお、MeCab の辞書には新語や固有表現に強いとされる mecab-ipadic-neologd を使用している。ページの閲覧が終わるまで読まれた文章について形態素解析処理を行い、ページが変わるなどユーザーが閲覧を終えることで、ページ閲覧が終了した後に実行される処理に移る。ページ閲覧が終了した後に実行される処理では保管した形態素解析結果から全体の解析をトピック分析などで使用されている tf-idf という出現頻度に基づく評価手法を用いて行う。この tf-idf

は、図 3 で示した (1), (2), (3) の式を用いて、ある単語 (x) が文章中に何回表れるか、またその単語が文章群においてどれほど稀な出現をしているのかという 2 種類の値を組み合わせて文章が含む単語の重要度を推定する。これにより、文章の特徴的な要素を判別することができ、膨大にある単語のかたまりから情報の把握に必要な単語がどれかであるか推測可能となる。ただし「は」、「の」、「です」のような頻出の単語であるが検索語としては採用されない単語については、ストップワードとして解析対象外としてある。本システムでは、この処理によって推測された重要語のうち重要度の高い上位 10 単語が保存されて、次の内容把握フェーズでの解析に移る。

(1) tf (Term Frequency)

$$= \frac{\text{文章Aにおける単語Xの出現頻度}}{\text{文章Aにおける全単語の出現頻度の和}}$$

(2) idf (Inverse Document Frequency)

$$= \log \left(\frac{\text{全文章の数}}{\text{単語Xを含む文章数}} \right)$$

(3) tf-idf = tf × idf

図 3 tf-idf スコアの算出式

4.3 内容把握フェーズ

前フェーズで使用した tf-idf は単語の出現頻度のみを考慮した指標であるため、単語の意味情報が考慮されていない。そこで、当フェーズでは実際に単語が持つ意味的な情報を、単語の空間ベクトル表現を用いて、単語間の関係から評価していく。今回は単語のベクトル化手法の 1 つである word2vec[10]を用いて、単語の空間ベクトルモデルを作成し、このモデルを元に最終的にユーザーに呈示される検索クエリを決定する。word2vec とは、同じ文脈に出現する単語は近い意味を持つという仮定に基づき、単語の大量のテキストデータを文章における単語の距離や対象とする語 X に単語 A が適合するかなどを解析、学習することで各単語の意味をベクトルで表現することが可能な単語のベクトル化手法である。このように複数の単語を空間ベクトル化したモデルである分散表現を作成して使用することで単語の加算や減算、類似度の評価など様々な処理が可能となる。モデル作成のためには大量のテキストが必要となるため、Wikipedia が公開している全記事の本文テキストデータ(約 3GB)を使用した。学習のためのコーパスは Wikipedia 記事を MeCab で分かち書きして作成しており、モデルの次元数は 200 次元で作成した。今回は文章解析フェーズより送られてきた単語 10 語に対して、word2vec を使用して単語の類似度を求めていき、情報がベクトル的に近く、一方向にそろったものを判別する。その後、判別された単語ベクトルの近い 5 単語をまとめて検索クエリとして、ユーザーにポップアップ通知を行い、ユーザーはこのポップアップ通知をクリックすることで、作成された検索クエリでの Google 検索が自動的に行われ、結果が表示される。これによりユーザーが閲覧した web ページに関連する追加情報の提示が可能となる。

5. システムの精度検証

いくつかの web ページで、本システムによって閲覧ページのコア情報を検索クエリとして提示できるかについて検証を行った。本稿では分析結果の一つとして北海道に存在する特別史跡である五稜郭のホームページ(図 4)を閲覧した際の結果について示す。視線位置に存在した文に形態素解析を行い、ストップワードを除去した上で tf-idf の計算を行った。その結果、重要語として、重要度が高い順に「五稜郭、指定、龍岡城、西洋式、城郭、庁舎、星、利用、特別史跡、堀」の 10 語を得た。さらに word2vec で単語間の類似度比較の結果、「龍岡城、城郭、特別史跡、指定、庁舎」の 5 語を得た。今回は word2vec の結果で得られた上記の 5 語を検索クエリとしてシステムに反映させた。その結果、日本に存在する特別史跡である熊本城跡に関する情報、同じく五稜郭である龍岡城五稜郭に関する情報を関連情報として被験者に提供することができた。前者は特別史跡に関する関連情報、後者は五稜郭に関する関連情報として提示された。よって、視線情報を利用して、閲覧するだけで関連情報を提供することができたとと言える。しかし、今回検証に使用した web ページの一部には広告やおすすめ記事へのリンクなど、ページの主な内容とは異なった文章が存在しているページがあり、これらの関係のない文を閲覧することで、重要語の解析及び、単語間の類似度比較に悪影響を及ぼしていた。よってタイトルと本文のみを解析対象とする等、本文の内容に関係のない文章は解析しないように改善する必要があると考える。

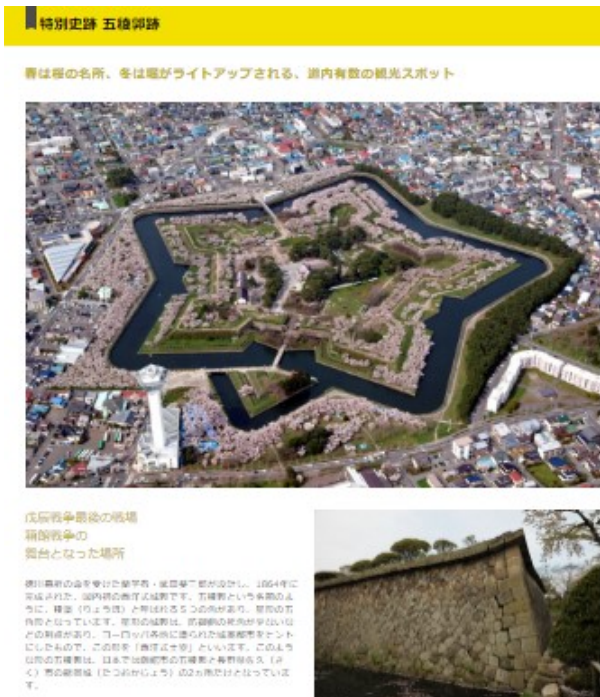


図 4. システム検証に使用した web ページ

次に、本システムによって導き出される検索クエリの精度が異なるページ間で差が生じるかについて、東海大学の学生の男女 15 名の協力のもと、アンケート調査によって、システムの評価を行った。アンケート内容は北海道にある五稜郭のホームページとキツザニアの紹介記事で本システムによるキーワード抽出を行い、最終的に提示される 5 単

語についてそれぞれのページの内容に沿った単語であるか、A: 適している B: やや適している C: 適していないの 3 段階で評価を行った。

また、今回被験者に提示した検索クエリは全員がページの本文をすべて閲覧したと仮定して解析を行っているため、被験者ごとの未読文の有無による違いは考慮されていない。本システムの定量的な評価には Average Precision(以下、AP とする)を用いた。AP は画像識別や物体検出に有効な評価指標であることが知られており、推薦システムにおいても推薦精度を定量的に評価することが可能である。図 5 に AP の算出式を示す。R は適合アイテムの数、r はアイテムの推薦順位、 $Prec(r)$ は適合アイテムの集合を A、システムが推薦したアイテムの集合を B とした場合の適合率で、 $Prec = \frac{|A \cap B|}{|B|}$ によって求められる。I(r) は推薦順位 r 位の推薦アイテムが適合アイテムの場合 1、そうでない場合は 0 となる関数である。これにより、推薦アイテムの内適合したアイテムが見つかる度に、その位置での適合率 $Prec(r)$ の計算を行う。

今回は本システムによって提示された 5 単語を推薦アイテム、アンケートの回答の内 A: 適していると回答された単語のみを適合アイテムとしてユーザーごとに AP の算出を行い、最終的に全ユーザーの AP の平均を取った Mean Average Precision(以下、MAP とする)によって評価を行った。また、アイテムの推薦順位については tf-idf による解析によって導き出された重要度を降順にしたものを使用した。評価結果を表 1 に示す。五稜郭のホームページから抽出した 5 単語の MAP の値は 0.34、対してキツザニアの紹介記事から抽出した 5 単語の MAP の値は 0.74 となった。これらの結果に有意な差があるかについて、ウィルコクソンの符号順位検定を行った。有意水準 $\alpha=0.01$ として検定を行った結果、 $p=0.007 < 0.01$ となり、両者には有意差が認められた。

表 1 ユーザごとの AP スコア

ユーザー	AP	
	キツザニア	五稜郭
A	1	0.59
B	1	0.37
C	0.45	0.2
D	1	0.45
E	0.45	0.33
F	0.87	0.42
G	0.87	0.33
H	0.87	0.37
I	0.89	0
J	0.45	0.37
K	0.59	0.37
L	0.53	0.37
M	0.53	0.37
N	0.87	0.37
O	0.76	0.2
MAP	0.74	0.34

両者の結果にこれほどの差が出た理由について、ホームページが表している内容が一意的であるか否かということが一つの要因として挙げられると考察する。例えば精度の高かったキッサニアのホームページでは「キッサニアとはこういう場所でこんなことができる」と伝えたい内容が簡潔かつシンプルであったが、五稜郭のホームページでは建物の紹介の他にも、建造された歴史なども含まれており、ページの伝えたい内容が一意的なものではなかったため、ユーザーによってどこを重要と捉えるかが異なり、精度がやや低下したと考察する。

6. システムの処理時間の検証

次に、このシステムの処理時間の検証を実施した。本検証は、東海大学のニュースページから 10 ページ無作為に選択し行った。本システムは、ユーザーの視線に基づいて文を抽出及び分析し、ユーザーを別の Web ページに誘導するための検索クエリを作成する。したがって、処理時間は、ユーザーがページの閲覧を終了してから検索クエリのポップアップ通知が画面に表示されるまでの時間とする。結果、今回の検証では、10 ページの平均処理時間は約 4.0 秒であった。Nielsen Norman Group[11]の調査によると、インタラクションシステムの応答時間について、ユーザーの思考の流れを止めない限度は 1 秒以内であり、それを超えると徐々にユーザーの意識が別のことに向くという報告がされている。このことから、我々のシステムの処理時間はインタラクションシステムとして不十分であるといえる。tf-idf の計算及び、word2vec での評価が処理時間に影響を与えていると考えられるが、特に word2vec モデルを使用した評価は、モデルのロードに時間がかかるため、処理時間が長くなる主な原因であると考えられる。またページ間の文字数の差が処理時間に与える影響について、今回の検証では傾向は確認できなかったため、より文字数差の多いページ同士の比較をしていく必要がある。

表 2 ページごとのシステムの処理時間

ページ	処理時間(秒)	文字数
A	3.99	1824
B	3.89	582
C	3.98	1025
D	3.85	765
E	3.87	1257
F	3.94	1099
G	3.98	855
H	4.01	866
I	3.99	1648
J	4.03	800
avg.	3.95	1072

7. おわりに

ユーザーの視線に基づいて文章を抽出し、重要語及び単語間の類似度の高い単語群を検索クエリとして提示を行う、関連情報提示インタラクションシステムの開発を行った。実際の web ページ閲覧によるシステム検証によって注視位置に存在する文を抽出し、ページのコア情報をユーザーインタラクションに使用できることを確認した。またアンケートによる精度評価から閲覧するページによってシステムの精度が大幅に変わるという問題点も発覚した。さらにシステムの処理時間についてもインタラクションシステムとして不十分であると判明した。今後は、文章抽出、解析の改善や、注視時間での重みづけによる利用者の興味の反映することを旨とする。

謝辞

本研究の一部は科研費(20K12411,17K02129)の助成を受けた。

参考文献

- [1] 総務省:令和 2 年 情報通信白書 ICT サービスの利用動向
- [2] 渡辺奈夕子, 岡本昌之, 菊池匡晃, 飯田貴之, 佐々木健太, 堀内健介, 大村寿美, 服部正典, “閲覧 Web ページからのキーワード抽出に基づくモバイル端末向け検索支援 UI”, 情報処理学会インタラクション 2011(2011)
- [3] 岡本昌之, 渡辺奈夕子, 大村寿美, “タッチ操作によるモバイル機器向け情報検索支援”, 東芝レビュー, Vol.66, p46-49(2011)
- [4] 上村卓史, 喜田拓也, 有村博紀, “ウェブ閲覧における効率的なキーワード抽出とその利用”, 情報処理学会論文誌, Vol. 1, No. 1, p49-60(2008)
- [5] tobii, “Specifications for the Tobii Eye Tracker 4C”, <<https://help.tobii.com/hc/en-us/articles/213414285-Specifications-for-the-Tobii-Eye-Tracker-4C>>, viewed 9 January 2021, (2020)
- [6] Taiga Mori, Takahide Otomo, Yusuke Nosaka, Eriko Ishii, Yuko Hoshino, Mitsuho Yamada, “Development of a Web Browsing Support System Using Eye Movements”, 2020 IEEE 9th Global Conference on Consumer Electronics, pp.714-718(2020)
- [7] 斎田真也, “速読と眼球運動”, 日本基礎心理学会, Vol. 23, No. 1, p64-69(2004)
- [8] 小林潤平, 川嶋稔夫, “日本語文章の読み速度の個人差をもたらす眼球運動”, 映像情報メディア学会, Vol. 72, No. 10, pp. J154~J159(2018).
- [9] 株式会社アイ・エム・ジェイ, “タブレット端末でのサイトユーザービリティ調査-画面サイズの違いによるタップのしやすさ, 読みやすさを検証-”(2013)
- [10] T. Mikolov, I. Sutskever, K. Chen, G. Corrad, J. Dean: “Distributed Representations of Words and Phrases and their Compositionality”, NIPS 2013: Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol.2, pp.3111-3119(2013)
- [11] Jakob Nielsen, “Website Response Times”(2010)