

# Web 検索時の思考プロセスの形式化に向けた解析手法の検討

## Analysis method for formalization of thought process at web search

永野 一馬<sup>†</sup>      荒川 豊<sup>†‡</sup>      安本 慶一<sup>†</sup>  
 Kazuma Nagano    Yutaka Arakawa    Keiichi Yasumoto

### 1. はじめに

今日、人々は多くの情報に晒されており、インターネット上に溢れる多種多様なコンテンツの中から、目的に適した情報を見つけ出さなくてはならない。ウェブ環境の情報検索におけるユーザ特徴を調査した先行研究 [1] では、高い情報検索能力を持つユーザ群に対して、高学歴、批判的思考能力が高いといった特徴が見られており、情報の検索効率と学習能力、労働生産性の関連が報告されている。一方で、ユーザの情報検索能力が低い場合、情報過多により探している情報を見つけ出せない場合や、精度の低い誤った情報が多くのユーザ間で拡散共有されてしまうといった問題がある。

また、これらの理由より、より高い情報検索能力を持つキュレータと呼ばれるユーザが様々なコンテンツを特定の目的のもとに編纂したキュレーションメディアが近年注目されている。こういったメディアを利用することで、ユーザは自身の能力に依らず、情報過多の状態及びそれに伴う負荷を軽減しつつ、目的の情報を得ることができる。しかし、キュレータが集めた情報をただ利用するユーザは、提供されるコンテンツがどういった思考プロセスを経て得られたものなのかを知ることができないため、自身の情報検索能力の向上を図ることができない。また、情報検索に関する先行研究では主に、1) クエリに対するコンテンツの適合性、2) ユーザの情報要求の推定、3) 検索システムの性能評価といったトピックに着目しており [2]、検索時のユーザの思考プロセスそのものの形式化に着目した研究は少ない。

Jessica[3]らが報告しているように、ユーザのブラウザの検索履歴には多くの思想、趣向といった情報が含まれており、思考プロセスを共有するために、そのまますべての履歴情報を共有することは非常に危険である。自身のブラウザの検索履歴の編集を試みても、膨大な URL リストの中から自身の思考プロセスを抽出し共有可能な知識に変換することは難しい。そのため、多くのユーザにとって情報検索時の思考プロセスは、後から振り返り、互いに評価することが困難な暗黙知となっている。

本研究では、情報検索研究における新たなトピックと

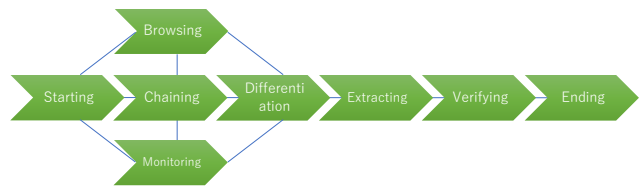


図 1: Ellis モデル概要図

して、暗黙知である Web コンテンツ検索時のユーザの思考プロセスをブラウザの検索履歴から抽出し、他者への経験サプリメントとして提供可能な知識に変換することを可能にする手法を検討する。本稿では、ブラウザ履歴の持つプライバシー情報への配慮より、ユーザの思考プロセスをユーザー自らが省みることができるようブラウザの検索履歴から抽出し、ユーザ自身の思考プロセス知識化する作業を補助するツールを提案する。

### 2. 関連研究

#### 2.1 情報検索時の思考プロセスの定義

偏にブラウザを用いた情報検索と言っても様々な形がある。文献 [4] では、まず、よく知られた情報検索プロセスとして、“単語の意味”や“名前がわかっている店の位置”といった検索対象に関する知識が十分にあり、明確な情報要求がある場合をあげている。このような検索は Look-up Search と呼ばれ、基本的には 1 つのユーザクエリに対して一つのコンテンツをシステムが返すことでシナリオを終える。この場合、内在する思考プロセスはわずかであり、効率的に情報にたどり着くためには、情報検索能力よりも事前の検索対象に対する知識が重要となる。一方、Web 上のコンテンツの多様性が広がるにつれ、「～の知識を身につけたい」、「新しい家電を購入したい」といった、より抽象度の高い、コンテンツと要求の適合判断が困難な、何度も検索を繰り返すタスクが行われるようになった。このような検索を Exploratory Search と呼び、こういった曖昧で移り変わりやすい情報欲求に対して、適合性の高いコンテンツを提供する研究が近年盛んにされている。Exploratory Search の定義として、先行研究であげられている特徴を文献 [5] がまとめている。またその中であげられた特徴にもっとも合致する既存の情報検索モデルとして Ellis モデル [6] を挙げている。Ellis モデルは研究者へのインタビューを通じて

<sup>†</sup> 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

<sup>‡</sup> JST さきがけ, Japan Science and Technology Agency.

情報探索活動を評価しており、8つの情報探索プロセスで構成されている。このモデルでは、Starting（情報検索の開始）プロセス後、Chaining（後方または前方 - 最初の情報源における参照によって続く）、Browsing（キーワード検索などの直接的な検索行動）、Differentiation（品質と関連性の判断に基づく情報源のフィルタリングと選択）、Monitoring（情報分野の発展状況の把握）のプロセスを繰り返し、Extracting（情報源から興味のある情報の系統的抽出）、Verifying（確度の確認）、Ending（最終検索、対象となるすべての情報を確認する）をもってユーザは検索行動を終了する。また、Ellisモデルでは上記のプロセスの時系列順序が変化することを特徴としており、その点も、Exploratory Search との適合性がみられる。本研究ではユーザの情報検索能力への関わりが深い Exploratory Search 時のユーザの思考プロセスに着目し、ユーザの思考プロセスモデルとして Ellis モデルを想定する。

## 2.2 既存の Web ブラウザ履歴の可視化手法

ユーザの Web ブラウザの閲覧履歴の可視化を試みた近年の研究として、Matthew らの popHistory と呼ばれるアプリケーションがある [7]。この研究では、アニメーションによりバブルチャートを時系列で表示し、ユーザの訪問回数によりバブルの大きさを変化させることで、自身のブラウジング履歴中の重要度の高いページをより簡単に発見できるようにすることを目指している。また従来のリストビュー形式の履歴と比較して参加者がポジティブな反応を示したことを報告しているが、このアプリケーションでは異なる検索目的のサイトを同一の1日のバブルビューとして表示しているため、ユーザのブラウジング回数が増えるにつれて自身の思考プロセスを省みることが難しくなってしまう点や、重要度の高いページでも、1度しか閲覧していない Web ページはアプリケーション上で発見することが困難になってしまうという問題点がある。

## 2.3 Web ページのユーザエンゲージメント評価

本研究では、Web サイトのユーザの重要度の評価手法として、Legan らの提案する ViewPort 時間を用いた評価手法 [8] を用いる。彼らは yahoo に掲載されている 1,971 のニュース記事に対する 267,210 のページビューに対するユーザの閲覧パターンを調査した。彼らはユーザの各ニュース記事の滞在時間のみでは Web ページのどの程度までユーザが閲覧したかがわからない点を指摘し、表示領域ごとの表示時間を Viewport 時間として定義し、ユーザの閲覧行動パターンをマルコフモデルを用いて分析した。ユーザがより多く Web ページを読むほどエンゲージメントが高いという仮説を立て、ViewPort 時間

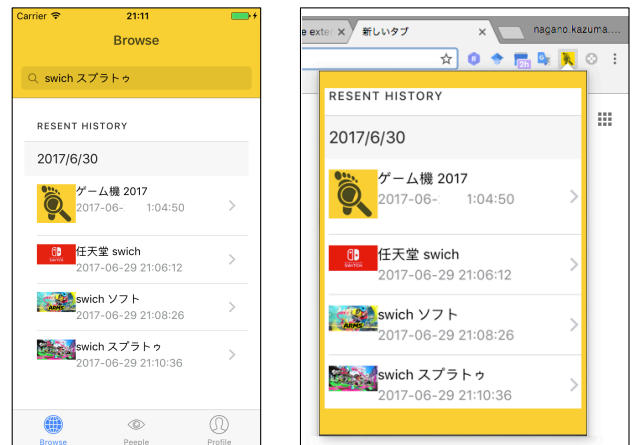


図 2: データ収集アプリケーション

をもとにユーザが“すぐにページを離れる”，“すべての記事を見た上でトップ位置までスクロールで戻る”等の記事に対する行動パターンを予測し、ユーザの ViewPort 時間より下記の4つのエンゲージメントレベルを用いた評価手法を提案した。

- Bounce: ページを訪問後すぐに離脱 (滞在時間が10秒以内)
- Shallow engagement: 記事本体の閲覧\* が50%以下。
- Deep engagement: 記事本体の閲覧が50%以上
- Complete engagement: 記事のコメント欄でコメント記載、または返信

先行研究では Complete Engagement のみ Bi-model の形で定義されているが、本研究では対象がニュース記事にとどまらないため、別の評価手法を定義する必要がある。そこで、本研究では、ユーザの Exploratory Search の目的を有用な情報を探し出すこととし、最も高いエンゲージメントを示した行動を、

- Complete engagement: ユーザがタブ切替により再びページを訪れた、または Bookmark に登録した。

として再定義し、Web ページの重要度の評価手法として用いる。

## 3. 提案システムの概要

本稿では、ユーザの思考プロセスをブラウザの閲覧履歴から抽出し、ユーザ自身の思考プロセス知識化作業を補助するためのツールを検討するため、1) ユーザの閲覧時の行動データの収集方法、2) 収集されたデータの解析方法の2つを提案する。

\* 5秒以上の Viewport 時間を閲覧と定義する。

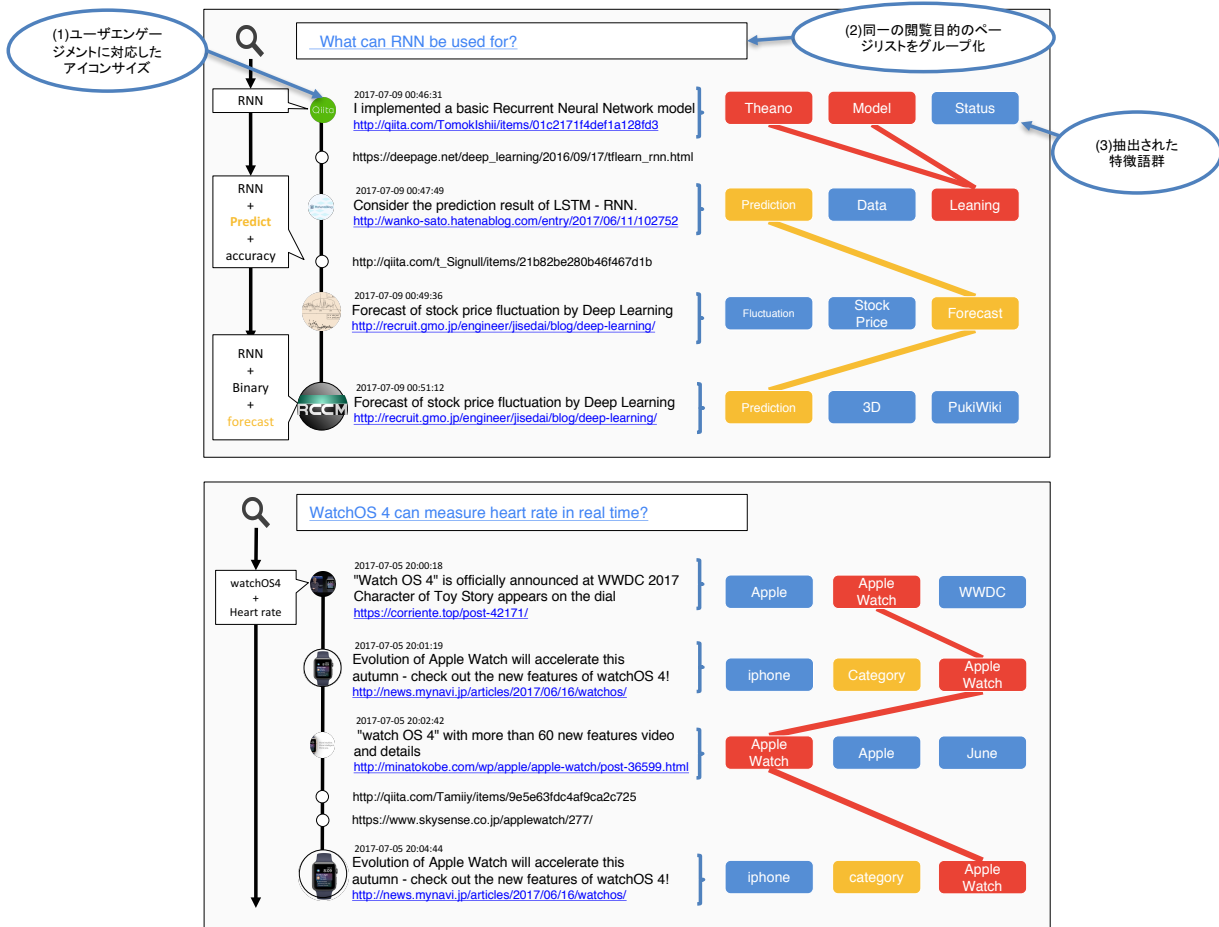


図 3: 編集された思考プロセス画面

### 3.1 データの収集方法

近年のスマートフォンによるモバイルネットワークの普及に伴い、ユーザの Exploratory Search 時の環境として、スマートフォンを用いて気軽に短時間検索を行う場合や、PC を用いて長時間集中して検索を行う場合が考えられる。そこで、本研究ではブラウザ型のスマートフォンアプリケーションと、Google Chrome 拡張機能の図 2 (左:スマートフォン, 右:Chrome 拡張機能) に示す 2 種類のアプリケーションを開発した。なお、スマートフォンアプリケーションは iOS と Android, GoogleChrome 拡張機能は Windows と Mac OS に対応している。ユーザは同一のアカウントでログインすることで、別環境で収集された閲覧履歴を共有することができる。また、思考プロセス抽出のためのデータとして、ページの閲覧履歴、Viewport 時間、スクロール速度、タブ切り替えイベント、Bookmark イベント、位置情報といった多種多様なデータを収集することができる。本稿ではこれらのうち、閲覧履歴、検索キーワード、Viewport 時間のみを評価で用いるが、収集されるその他のデータも今後思考プロセス抽出のための解析対象となる。

### 3.2 思考プロセスの抽出と可視化

上記のアプリケーションにより収集されたデータから以下の情報を抽出する。

#### (1) ページのユーザエンゲージメント

文献 [8] の手法をもとに評価する。4 つの評価レベルをもとに、ページのメタデータから取得したサムネイル画像の大きさを 4 段階で表示する。また Bounce のページに関しては有用な情報が含まれていなかったページと推測できるため、URL のみを画面表示する。

#### (2) ページの閲覧目的

ユーザの Exploratory Search のスタートポイントとして、検索エンジンを用いたキーワード検索の開始を抽出する。エンドポイントとして以下のものを想定し、エンドポイントが発生するまでに訪問したページを同一の閲覧目的としてグルーピングする。1) 検索エンジンを用いた新たなキーワード検索, 2) ブックマークされたページへのアクセス, 3) 外部情報に基づいたアドレスバーへの直接入力, また、ユーザはツール上で異なるグルー

ブに分類されたページ群を一つに集約することができる。グループ名は、ユーザがコンテンツ閲覧の意図を表すタイトルを付けるものとする。

### (3) ページの特徴語の抽出

本稿では初期手法として、最も一般的な特徴語抽出手法の1つである TF-IDF を用いる。Document Frequency は日本語の Wikipedia の記事を文書とみなして求めた。

以上の手法によりデータ収集アプリケーションが収集したデータより抽出された情報をもとに実際に作成したサンプル画面を図3に示す。抽出された特徴語は、右から重要度の降順に並んでいる。また、ユーザの編集としては、同一目的であったページ群の集約、ページ群のタイトルの編集、特徴語を色付け、関連性の高い特徴語を線で結ぶといった作業を想定している。図3上では、抽出された特徴語をつぎの検索キーワードに用いており、ユーザ行動に、Ellis モデルにおける Chaining, Browsing, Differentiation, Monitoring のループが見られる。また、ユーザの思考として、WatchOS 4 の情報を Exploratory Search していく中で“AppleWatch” そのものの情報より“iphone”との連携部分に注目していたことを特徴語のランク遷移から読み取れる。

## 4. まとめ

本稿では、Web 検索におけるユーザの思考プロセスを追跡し、収集したデータを視覚化する初期手法を提示した。提案されたアプリケーションを用いて、滞留時間、スクロール量とその速度、ページ間のコンテンツ内容の遷移を収集し、各ページのユーザの重要度とページの特徴語の推移を抽出する。また、これらの情報を単一のユーザが編集できるビューにグループ化して視覚化することで、ユーザは閲覧履歴を容易に思考プロセスを変換することができるアプリケーションを提案する。今後の研究として、本稿の手法を基準とし、多様なユーザの思考プロセスの抽出手法を比較し、よりユーザの思考プロセスを反映し、ユーザ自身の編集の労力を最小化する手法を明らかにする。さらに、これらを他者に提供することができるアプリケーションを作成し、ユーザにパーソナライズされた他者の思考プロセスを適切なタイミングで提案するフレームワークを構築することで、思考プロセスの共有化がユーザの情報検索能力の向上することができるサプリメントとして活用できる可能性を検証していく。

## 謝辞

本研究は、JST CREST (JPMJCR16E1) と NAIST-CICP (想像力と国際協力を育む情報科学教育コア「プロジェクト型研究」) の助成によって行われたものである。ここに記して謝意を表す。

## 参考文献

- [1] 南友紀子, 岩瀬梓, 宮田洋輔, 石田栄美, 上田修一, 倉田敬子. ウェブ環境における情報検索スキル. 日本図書館情報学会誌, Vol. 62, No. 3, pp. 163–180, 2016.
- [2] W Bruce Croft, Donald Metzler, and Trevor Strohman. *Search engines: Information retrieval in practice*, Vol. 283. Addison-Wesley Reading, 2010.
- [3] Jessica Su, Ansh Shukla, Sharad Goel, and Arvind Narayanan. De-anonymizing web browsing data with social networks. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 1261–1269. International World Wide Web Conferences Steering Committee, 2017.
- [4] Ryen W White and Resa A Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, Vol. 1, No. 1, pp. 1–98, 2009.
- [5] Emilie Palagi, Fabien Gandon, Alain Giboin, and Raphaël Troncy. A survey of definitions and models of exploratory search. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, pp. 3–8. ACM, 2017.
- [6] David Ellis, Deborah Cox, and Katherine Hall. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation*, Vol. 49, No. 4, pp. 356–369, 1993.
- [7] Matthew Carrasco, Eunye Koh, and Sana Malik. Pophistory: Animated visualization of personal web browsing history. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2429–2436. ACM, 2017.
- [8] Dmitry Lagun and Mounia Lalmas. Understanding user attention and engagement in online news reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 113–122. ACM, 2016.