

D-014

Web 閲覧履歴におけるうろ覚え Web ページの対話的再発見 Interactive rediscovery of vaguely-remembered web page from web browsing history

井倉 真一[†] 近藤 司[‡] 伊藤 真也[‡] 原田 史子[†] 島川 博光[†]
Shinichi Ikura Tsukasa Kondo Masaya Ito Fumiko Harada Hiromitsu Shimakawa

1. はじめに

近年, WWW の普及により, ユーザは Web ページから手軽に情報を得ることができるようになった. 日々, Web ページを閲覧する中で, 過去に閲覧した, ある Web ページの情報が必要になり, 再訪問したいと思うことがある. しかし, もう一度閲覧したい Web ページに再訪問できない場合がある. 再訪問できない理由として, もう一度閲覧したい Web ページを発見するためのキーワードが思い浮かばないことが考えられる. ユーザが閲覧した時点では, 必要でない情報と考えていたので, キーワードが印象に残らなかったということも考えられる. 本論文では, 過去閲覧した Web ページを再発見するためのキーワードが分からない場合にも, 目的の Web ページに辿り着くための手法を提案する.

2. 閲覧履歴における検索

2.1 うろ覚え状態の Web ページ再発見

大量の Web ページを閲覧する過程で, 過去に閲覧した Web ページを再度, 閲覧したくなることがある. この場合, 検索エンジンを用いて, 再発見する必要がある. しかし, 検索エンジンを用いて, 再発見するには困難な場合がある. 例えば, 過去に "1 人暮らし用の料理レシピ 100 選" という Web ページを閲覧したとする. ダイエットに関する Web ページを探しているときに, 他の Web ページのリンクからその Web ページに偶然辿りついた. その時は実家住まいで, 自炊する機会が無かったため, その Web ページに興味を湧かず, 内容を詳しく読まなかった. その後, 一人暮らしをすることになり, 当該ページを再度, 閲覧したくなったとする. しかし, Web ページの内容を詳しく読んでいないため, 調理の動画があったことは何となく覚えているが, タイトルや具体的なレシピ名は覚えていない. そのため, 再発見するための適切なキーワードを思い浮かぶことは困難である.

このように, Web ページの内容を曖昧にしか覚えていない場合, 再発見するための適切なキーワードを思い浮かべず, 検索エンジンを用いることができない. このような状態をうろ覚え状態と定義する. もう一度閲覧したい Web ページを目的 Web ページと定義する.

2.2 Web 閲覧と結びつく情報

ユーザは Web ページを閲覧したときに, 閲覧したという事実, あるいは当該ページの漠然とした内容を, ささまざまな情報から記憶している. 例えば, 2.1 節の例の場合に, 閲覧した "1 人暮らし用の料理レシピ 100 選" に対して, "3ヶ月ほど前に閲覧した", "他の Web ページのリンクから辿り着いた", "ダイエットに関する Web ページを探している時に偶然, 閲覧した", "実際に調理している動画があった" などが挙げられる. これらの情報は, 閲覧履歴から目的 Web ページを探す手がかり

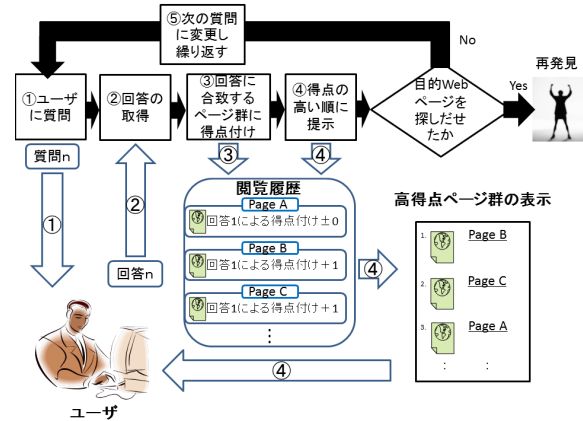


図 1: 全体図

になる. "3ヶ月ほど前に閲覧した" という情報に合致する Web ページは, 閲覧履歴の中に多く含まれていると考えられるが, 他の 3 つの情報と組み合わせることで目的 Web ページを閲覧履歴から絞り込めると考えられる. Web ページの閲覧と結びつく情報を条件として閲覧履歴を検索することで, うろ覚え状態でも, 目的 Web ページを探し出せると考えられる.

3. うろ覚え状態における目的 Web ページの再発見支援

3.1 対話的質疑による目的 Web ページの再発見

本論文では, 2.2 節で述べた情報を用いて, ユーザのうろ覚え状態における目的 Web ページの再発見を支援する手法を提案する. 本手法では, ユーザに対話的に質問し, その回答を取得することを繰り返す. 手法の全体図は図 1 に示す通りである. まず, ユーザへ質問し, それに対する回答を取得する. そして, ユーザの回答に合致する Web ページへの得点付けをし, 閲覧履歴から得点の高い順に Web ページ群を提示する. 最後に, ユーザに対し, 質問と回答をもう一度繰り返すかどうかをユーザに提示する.

3.2 質問のための指標群

ユーザに対する質問の回答から目的 Web ページの特徴を把握することで, 閲覧履歴からユーザの目的 Web ページを絞り込む. ここで, 閲覧履歴から目的 Web ページの候補を絞り込める情報を指標と定義する.

2.2 節で述べたように, ユーザが目的 Web ページを閲覧した事実, あるいは漠然とした内容を表す情報が探す手がかりになると考えられる. 本論文では, 「閲覧時期」「ルート」「閲覧目的」「メディア情報」の 4 つに関して指標を設定する.

指標は, 以下の 2 つの条件を満たす必要がある. 1 つ目としては, うろ覚え状態においても, ある程度ユーザが記憶している内容であるということである. 2 つ目としては, 閲覧履歴から目的 Web ページの候補を絞り込

[†]立命館大学情報理工学部

[‡]立命館大学大学院理工学研究科

むことができるということである。この2点に関する各指標の有用性について説明する。

過去のユーザの目的 Web ページの閲覧が、現在からどれくらいの期間以内であったかを閲覧時期と定義する。過去の Web ページの閲覧を思い出す場合、いつ頃前に閲覧したかということを考えることがよくある。いつ頃前に目的 Web ページを閲覧したかを指標に用いることで、閲覧履歴から目的 Web ページの候補を絞り込む事ができると考えられる。

ユーザが目的 Web ページへどのような方法で辿り着いたかを表す情報を、ルートと定義する。文献 [1] に基づいてユーザが Web ページに辿り着く方法を分類すると、検索エンジンを用いたか、Web ページのリンクからか、その他かの3つが考えられる。以下の2つに関して質問をすることで、目的 Web ページへのルートを把握することができる。1つ目は、検索エンジンを利用したかどうかという質問をする。2つ目は、リンクから辿り着いたかどうかという質問をする。閲覧履歴の各ページのルートとこれらと比較することで、閲覧履歴から目的 Web ページを絞り込める。

ユーザが Web ページを閲覧する上での目的を閲覧目的と定義する。人は閲覧目的をもって Web ページを閲覧する。閲覧目的を達成するために、複数の Web ページ群を閲覧する。2.1 節の例では、ダイエットに関する情報を得るといった閲覧目的を持っている時に、目的 Web ページを閲覧した。閲覧履歴をユーザの閲覧目的ごとにクラスタに分類できることが分かっている [2]。よって、閲覧目的は指標として有用と考えられる。

Web ページにおける、動画や画像の有無をメディア情報と定義する。Web ページを閲覧するさいに、視覚的な特徴がある場合、その特徴が印象に残りやすいと考えられる。以下の2つの項目により、メディア情報を把握する。1つ目は動画が存在するかどうかということである。2つ目はあるサイズの画像が存在するかどうかということである。Web ページに存在する画像には、さまざまなサイズが存在する。大きなサイズの画像は印象に残りやすいと考えられる。そこで、ある大きさ以上の画像が含まれていたか否かを考慮することで、比較的印象に残りやすい画像の情報を指標として扱える。

ユーザがメディア情報を含む Web ページをよく閲覧するか否かによって、閲覧履歴における、メディア情報を含む Web ページの母体数が変化する。そのため、メディア情報に関する質問をしても、目的 Web ページの候補を絞り込めないことも考えられる。しかし、前述のとおりメディア情報はユーザにとって、印象的で間違え難い指標であると考えられる。そのため、メディア情報は目的 Web ページの候補を絞り込むには難しいが、印象的な指標であるため、より信頼性が高いと考えられる。よって、メディア情報は有効な指標である。

本手法はこれら4つの指標に基づき、質問を生成する。

3.3 質問に対する回答と得点付け

3.2 節の指標を用いた質問に対して、ユーザは、目的 Web ページが指標に合致する場合は「はい」、合致しない場合は「いいえ」、回答不能な場合は「分からない」を回答する。ユーザはうろ覚え状態であるため、目的 Web ページの特徴に対して詳細な情報を入力できないと考え

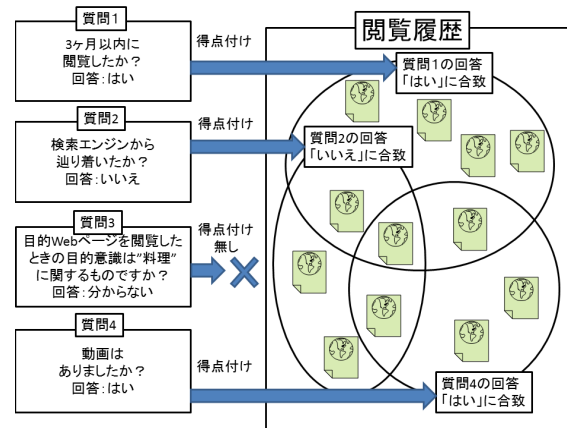


図2: 回答と得点付け

られる。そこで「はい」、または「いいえ」で回答させることで、ユーザは目的 Web ページに質問の指標の特徴が含まれるか否か判断するだけで良い。

ユーザの回答に基づき、Web ページの得点付けをする。質問の数だけ、得点付けを繰り返す。図2は閲覧履歴の Web ページ群に対する得点付けの図である。質問に対するユーザの回答結果に合致する Web ページのみを得点付けする。「分からない」と回答された質問に対してはどの Web ページにも得点付けしない。

3.4 目的 Web ページ候補の提示と得点付けの繰り返し

質問とユーザの回答による得点付けの後に、得点の大きい順にランキング形式で Web ページを提示する。これにより、目的 Web ページの可能性の高い Web ページを優先的に、ユーザに提示できる。さらに、質問群に対する回答結果群により合致する Web ページであれば、ユーザの目的 Web ページである可能性が高い。そのため、質問が繰り返される度に、得点の高い Web ページに目的 Web ページが含まれる可能性が高くなる。そこで、1つの質問とユーザの回答による得点付けが終了する度に、ユーザに繰り返すかどうかを選択させる。ユーザが繰り返すを選択した場合、もう一度質問と回答による得点付けを繰り返す。提示する Web ページを質問ごとに更新することで、ユーザはその都度、提示される Web ページの中に目的 Web ページがあるかどうかを確認できる。

4. おわりに

本論文では、過去に閲覧したうろ覚え Web ページの再発見の支援方法を提案した。今後は指標の有用性を検証していく予定である。

参考文献

- [1] 第26回 インターネット利用者アンケート結果: <http://docs.yahoo.co.jp/info/research/wua/201012/page14.html>
- [2] K Higuchi, F Harada, H Shimakawa, Making Links between Goal and Tasks to Retrieve Information from Web Browsing Histories, International Conference on Computer Sciences and Convergence Information Technology pp.992-994,2009.