

Web 検索における情報見落としを防ぐ Web ページの提示手法の提案 A Study for Preventing Oversight of Information in Web Search

井倉 真一[†]
Shinichi Ikura

原田 史子[‡]
Fumiko Harada

島川 博光[‡]
Hiromitsu Shimakawa

1. はじめに

WWW ユーザは必要な Web ページを得るために、Web ページを一瞥することでその Web ページを詳しく閲覧するか否かを判断する傾向がある。理由として、WWW 上の莫大な Web ページ群の中から少しでも効率よく情報を探し出したいからであると推測できる。その結果、ユーザが一瞥し閲覧しないと判断した Web ページに、実は必要な情報が載っていたとしても見落とし、正しく情報探索が行えていない可能性がある。本論文では、WWW ユーザが Web ページの見た目だけで詳しく閲覧するか否かを判断することをできるだけ回避させる Web ページの提示手法を提案する。ユーザが Web ページを詳しく閲覧しないと判断するに至った Web ページの見た目上の要素の特徴を閲覧履歴から抽出する。その特徴を含む Web ページを閲覧するさいには、その要素の特徴部分を改変してユーザに提示することで、情報探索が正しく行えるように支援する。

2. Web 検索の現状

2.1 ユーザの検索傾向

ユーザが Web 検索をするさい、閲覧しようとしている Web ページで必要な情報を得るのが容易そうかを一瞥して判断し、詳しく閲覧するか否かを判断する [1]。これによって、検索効率を上げていることが考えられる。ユーザが必要な情報を得るのが容易そうか否かを判断する要因として、見た目が考えられる。ただしスクロールを必要とする部分のユーザの興味は薄いことが既存研究として分かっている [2]。したがってユーザは、スクロールなしに見える部分の見た目を中心に、詳しく閲覧するか否かを判断する。しかし、このようなユーザの閲覧傾向によって、必要な情報の見落としが発生することが考えられる。例えば、あるユーザは京都で部屋を借りたいと考えているとする。Web 検索エンジンの検索結果で、ある賃貸会社の Web サイト A と別の賃貸会社の Web サイト B が得られたとする。まずユーザは Web サイト A に訪れた。しかし、一瞥して賃貸サイト A は画面当たりの文字量が多く使いにくそうと感じたため、Web サイト A は利用しないと判断した。ユーザは他の賃貸サイト B で部屋を探すことにした。しかし実は、Web サイト A にはユーザの条件に一番合致する部屋が掲載されていた。その結果、ユーザは条件に一番合致する部屋を見つけることができなかった。このように、実はユーザにとって必要な情報があっても関わらず、見落とししてしまうことが考えられる。

2.2 Web ページの見た目における嫌悪要素特徴

ユーザには好ましくないと思う見た目に関する要素が存在することが考えられる。2.1 節の例の場合では、ユーザは文字量が多い Web ページを好ましくないと思っていた。つまり「文字量」という要素に対して、「文字量が

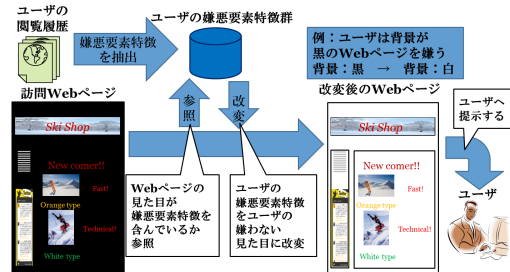


図 1: 全体図

多い」という特徴を嫌悪していて、それが詳しく閲覧しないという判断の理由になっているといえる。ここで、ユーザが訪問した Web ページを詳しく閲覧しないと判断する理由となった、Web ページの見た目上の要素の特徴を、嫌悪要素特徴と定義する。嫌悪要素特徴を持つ Web ページを、ユーザが嫌悪しない要素特徴に変更して表示することができれば、前節で述べた必要な情報の見落としは発生しないといえる。ただしユーザの知識や経験によって、Web ページの閲覧の仕方は異なる [3]。そこで本論文では、ユーザ毎の嫌悪要素特徴を Web 閲覧履歴から推定し、嫌悪要素特徴部分を非嫌悪の要素特徴に改変してユーザに提示するための手法を提案する。本論文では、見た目上の要素として「背景色」、「文字量」、「空白の割合」の 3 つを扱う。

3. 情報見落としを防ぐ Web ページの嫌悪要素特徴改変手法

3.1 手法の流れ

本論文では、2.2 節で述べたユーザの Web ページにおける嫌悪要素特徴を改変することで、ユーザの情報探索が適切におこなえるように支援する手法を提案する。本手法では、ユーザの閲覧履歴からユーザの Web ページにおける嫌悪要素特徴を特定し、ユーザが訪問を決定し、詳しく閲覧するか判断するであろう Web ページに嫌悪要素特徴が含まれる場合に、その要素特徴を部分的に改変してユーザに提示する。手法の全体図は図 1 に示す通りである。

まず、ユーザの閲覧履歴から嫌悪要素特徴を特定する。そして、ユーザがある Web ページに対する http リクエストを実行し、ブラウザ上で表示するとき、本手法により、その Web ページにユーザの嫌悪要素特徴が含まれているかどうかを判定する。嫌悪要素特徴を含まない場合は、Web ページを改変せずにユーザに Web ページを提示する。嫌悪要素特徴を含む場合は、対応する要素部分をユーザの嫌わない特徴に改変してユーザに提示する。ここで、改変後の特徴となるべき、ユーザの嫌わない要素特徴を、改変先要素特徴と定義する。

3.2 見た目嫌悪 Web ページ群の抽出

本手法ではユーザの嫌悪要素特徴を抽出する。そのためまず、ユーザの閲覧履歴の中から見た目が原因で詳しく閲覧しないと判断された Web ページ群を抽出する。

[†]立命館大学大学院情報理工学研究所

[‡]立命館大学情報理工学部

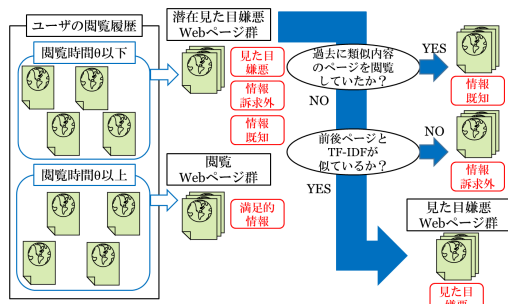


図 2: 見在目嫌悪 Web ページの切り分け手法

ここで、ユーザが Web ページの見た目が原因で詳しく閲覧しないと判断した Web ページを見在目嫌悪 Web ページと定義する。見在目嫌悪 Web ページは訪問直後に一瞥した結果、詳しく閲覧しないと判断され閉じられたため、閲覧時間は十分小さいと言える。しかし、閲覧時間が十分小さい Web ページが必ずしも見在目嫌悪 Web ページであるとは推定できない。したがって、見在目嫌悪 Web ページ群のみを正確に抽出するために、閲覧時間以外の性質も利用する。閲覧履歴内の各 Web ページをユーザが閉じたとき、当該 Web ページに対して何らかの評価心理を持っているといえる。ユーザが Web ページを閉じるときの心理として、以下の 4 つが考えられる。Web ページ内の情報に満足している心理状態を表す「満足的情報」。Web ページ内の情報が既知しているものであるという心理状態を表す「情報既知」。Web ページの訪問前は自分の欲しい情報を含んでいると期待していたが、実際に見るとそうではなかったという心理状態を表す「情報訴求外」。そして、見た目が原因でその Web ページを詳しく閲覧するのが煩わしいと感じる心理状態を表す「見在目嫌悪」。これに対応する Web ページが見在目嫌悪 Web ページである。閲覧履歴中から「見在目嫌悪」が理由で閉じられた Web ページ群を特定する。見在目嫌悪 Web ページ群の抽出手法を図 2 に示す。

まず、ユーザの閲覧履歴を閲覧時間で分類する。閲覧時間が 秒以下の Web ページ群は、「見在目嫌悪」「情報訴求外」「情報既知」が理由として閉じられた Web ページが存在する可能性が高い。このような Web ページ群を潜在見在目嫌悪 Web ページ群として定義する。また、閲覧時間が 秒以上の Web ページ群は、ユーザが満足して閉じた Web ページが存在する可能性が高い。このような Web ページ群を閲覧 Web ページ群として定義する。次に、潜在見在目嫌悪 Web ページ群の中から「見在目嫌悪」を理由として閉じられた Web ページだけを抽出する。そこでまず、「情報既知」を理由として閉じられた Web ページを除外する。そのために、潜在見在目嫌悪 Web ページ群の中の各 Web ページに対して、類似した内容の Web ページを過去に閲覧していたかを調べる。次に、残った潜在見在目嫌悪 Web ページ群から「情報訴求外」を理由として閉じられた Web ページ群を除外する。そのために、当該 Web ページの閲覧時刻が前後の Web ページとのコサイン類似度を調べる。最後に残った潜在見在目嫌悪 Web ページ群を見在目嫌悪 Web ページ群として抽出する。

3.3 ユーザの Web ページにおける嫌悪要素特徴の抽出

3.2 節で述べた手法によって閲覧履歴内の閲覧 Web ページ群と見在目嫌悪 Web ページ群を用いて、ユーザ

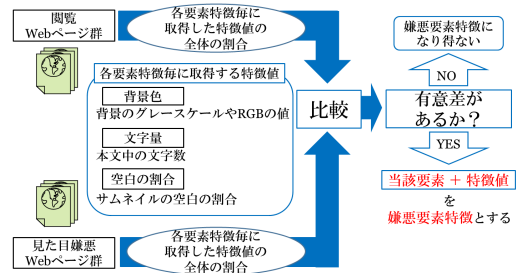


図 3: ユーザの Web ページにおける嫌悪要素の抽出法の Web ページにおける嫌悪要素特徴を抽出する。ユーザの Web ページにおける嫌悪要素の抽出法を図 3 に示す。

閲覧 Web ページ群と見在目嫌悪 Web ページ群に対して、見在目上の各嫌悪要素特徴毎に取得したデータの全体の割合をそれぞれ求める。今回は、2.2 節でも説明した通り「背景色」「文字量」「空白の割合」の 3 つを見在目上の要素として用いる。背景色はグレースケール値や RGB 値、文字量は Web ページ本文中の文字数、空白の割合はページのサムネイル画像において背景色が占める割合として数値化できる。この数値は対応する要素の特徴を表すので、特徴値と定義する。各要素の各特徴値に対して、当該特徴値に合致する特徴値を持つ Web ページ群の割合を、見在目嫌悪 Web ページ群と閲覧 Web ページ群からそれぞれ求める。求められた 2 つの割合に有意な差がみられるかを判定する。差が有意でない場合、当該特徴値は嫌悪要素特徴ではないと推定する。差が有意である場合、ある Web ページの見目の当該要素が当該特徴値であるということ嫌悪要素特徴として抽出する。

3.4 Web ページの嫌悪要素特徴の改変

嫌悪要素特徴が訪問 Web ページに含まれている場合、その特徴を改変先要素特徴に改変してユーザに提示する。改変先要素特徴は閲覧 Web ページ群で取得した特徴値の全体の割合をもとに決定する。例えば、閲覧 Web ページ群のうち、「空白の割合」要素について、特徴値「20%」である Web ページが半数を超えていた。この場合、「空白の割合が 20%」である Web ページはユーザによって閲覧される可能性が高いため、この特徴を改変先要素特徴として用いる。ユーザの Web ページの見目に対する嫌悪要素特徴が「空白の割合が 10%」とすると、もしユーザの訪問 Web ページにこの嫌悪要素特徴が含まれていた場合、「空白の割合が 20%」になるように Web ページを改変してユーザに提示する。

4. おわりに

本論文では、情報見落としを防ぐ Web ページの提示手法を提案した。今後は、嫌悪要素特徴の発見と、本手法の有用性について検証していく予定である。

参考文献

- [1] 種市淳子: エンドユーザーの Web 探索行動 Library and Information Science. v.55, 2006, p.1-23
- [2] Will Schroeder: Testing Web Sites with Eye-Tracking Sep 01, 1998
- [3] 斎藤ひとみ, 江草由佳, 高久雅生, 寺井仁, 三輪眞木子, 新門典子: Web 情報探索行動の分析: 課題の志向性と経験の違いによる影響についての予備的検討 第 13 回 Web インテリジェンスとインタラクション研究会, 2008-12