

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

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[†]

モバイル機器における Web 検索は一般的になりつつあるが、検索クエリの入力には PC の場合ほど容易ではない。本稿では、ユーザが Web ページを閲覧した際に、ページ中に含まれるクエリ候補のキーワードを抽出し、絞り込みをしてユーザに提示することで、ユーザが数回の選択操作だけで検索することを可能とする検索支援インタフェースを提案する。また 14 名の被験者によるオープン評価の結果を通じ、実用的な精度で利用できることを確認した。

Search Supporting UI for Mobile Terminal based on Keyword Extraction from Browsing Web Page

NAYUKO WATANABE,[†] MASAYUKI OKAMOTO,[†] MASAOKI KIKUCHI,^{††}
TAKAYUKI IIDA,^{††} KENTA SASAKI,[†] KENSUKE HORIUCHI,^{††}
SUMI OMURA^{††} and MASANORI HATTORI[†]

Inputting query terms on a mobile terminal is not easy though mobile web search is becoming popular. In this paper, we propose an information retrieval interface which enables users to search the web with only selecting operation by automatically extracts query-term candidates. According to our open evaluation with 14 users, our method achieved practical quality.

1. はじめに

スマートフォンなどのモバイル端末を用いた情報アクセスは増加しているが、Web 検索時の文字入力のコストは依然として課題である。このコストを減らすために、できるだけ少ない操作回数・操作時間で検索クエリを入力できる検索方式が望まれる。

ユーザが検索する可能性が高い単語をクエリ候補として推測・提示できれば、ユーザは選択操作のみで検索することが可能である。任意の状況でユーザが検索したい単語を推測するのは困難だが、現在閲覧中のページについて関連情報を調べる際には、検索クエリはそのページについての単語であることが多い⁵⁾。従って、閲覧中の Web ページ文章から検索クエリ候補となる単語を抽出することで、関連情報検索時に操作にかか

るコストを下げる事が可能であると考えられる。

我々はこれまでに、閲覧中の Web ページの関連情報検索を容易にする、キーワード抽出に基づく検索支援 UI を提案してきた⁶⁾。本稿では、これまでの知見に基づき改良した検索支援 UI を提案する。またスマートフォン上に実装した提案 UI の利用評価を行い、どの程度の割合で検索結果が利用されるかを調査した。

2. 検索支援 UI

本章では、モバイル環境においてクエリ入力の手間を削減しつつ Web 閲覧からの関連情報検索ニーズを満たす検索支援 UI の概要と、処理の流れを述べる。

2.1 利用シナリオ

提案 UI ではユーザが Web ページを閲覧した際、HTML に含まれるクエリ候補のキーワードを抽出し、絞り込みをしてユーザに提示する。このためユーザは以下のようにして、数回のタッチ操作のみで閲覧 Web ページの関連情報の検索が可能である (図 1)。

(1) Web ブラウザで Web ページを閲覧する。例え

[†] (株) 東芝 研究開発センター Corporate R&D Center, Toshiba Corporation

^{††} (株) 東芝 ビジュアルプロダクツ社 コアテクノロジーセンター Core Technology Center, Toshiba Corporation



図 1 検索支援インタフェースの利用例
Fig. 1 Example use of search-support application

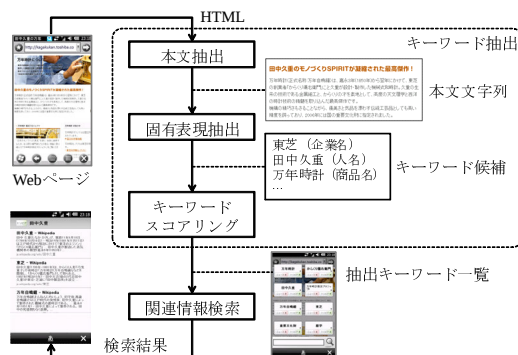


図 2 処理の流れ
Fig. 2 Process flow

- ば、万年時計についての記事を読んでいるとする。
- (2) 検索支援機能を利用可能な状態になると、画面上部に通知アイコンが表示される (図 1(a))。
 - (3) 閲覧ページに関連する情報を調べたい際は通知アイコンをタッチすると、ページ内の単語で検索に利用されやすいとシステムが判断したキーワード一覧が表示される (図 1(b))。キーワードのボタンは押す領域により通常のキーワード検索、ニュース検索、トリビア (Wikipedia) 検索が選択可能である。
 - (4) 万年時計の作者「田中久重」について詳細を知りたい場合、「田中久重」と書かれたボタンの「トリビア」部分を押すことで検索が行われ、検索結果が表示される (図 1(c))。
 - (5) 検索結果のスニペットをタッチすると対応する Web ページが表示される。

2.2 処理の流れ

以下では閲覧中の Web ページに対し、検索クエリに用いる可能性の高いキーワードを抽出し検索する流れを説明する (図 2)。提案 UI での処理は、キーワード抽出と関連情報検索のステップから構成される。

2.2.1 キーワード抽出

キーワード抽出は本文抽出、固有表現抽出、キーワードスコアリングの 3 つの処理から構成される。

本文抽出 本文抽出では HTML 文書を解析して本文領域を推定し、内部のテキストを抽出する。本文領域とはニュース記事などの Web ページのうち、広告部分などを除いた記事本文部分である。

HTML テキストをブロック要素毎に分割し、テキスト長、a タグで囲まれたアンカーテキスト長、句読点数、ストップワードなどを元にブロック毎の採否を決定する。採用するブロック内のテキストを出現順に連結したものが本文となる。

固有表現抽出 本文に対して固有表現抽出を適用し、固有表現クラスを付与する。固有表現クラスとは人物名、企業名などの固有名詞や、日時、数量など表層表現の意味を表す属性のラベルであり、辞書とその組み合わせルールによって付与される。例えば辞書に「松阪 (地名)」「牛 (食品)」が登録されている場合、登録されていない語「松坂牛」に対し、ルール「地名 + 食品 → 食品」を適用することで、固有表現クラス「食品」を付与する。このように固有表現クラスが付与された単語の一覧が、抽出キーワードの候補となる。

キーワードスコアリング キーワード候補に対し、スコアを付与する。スコアは、出現位置、文字列長、意味属性毎のスコア、文書内の出現頻度、WebIDF によるスコアの線形和として算出される。線形和の重みは、ユーザがどのようなキーワードを検索するかの事前調査に基づいて決定した。最終的に、スコア上位のものがユーザに提示されるキーワードとなる。以上のようにしてキーワード抽出が完了すると、画面に通知アイコンが表示され、ユーザが通知アイコンをタッチすると、1 画面あたり 8 個、最大 16 個まで、キーワードが表示される。

2.2.2 関連情報検索

提示されたキーワードの 1 つをユーザがタッチした場合、そのキーワードをクエリとして検索が行われる。ユーザがタッチする場所に応じ、1 つのキーワードについて検索ジャンルを切り替えられる。これはそれぞれのジャンルについて、検索クエリにサイト指定を追加する、特定の単語と AND 検索するといった処理を割り当てることで実現される。

2.3 実装

提案手法を、スマートフォン T-01A (OS: Windows Mobile[®] 6.5, CPU: 1GHz, 4.1 インチシングルタッチディスプレイ) 上のブラウザ (Internet Explorer[®]

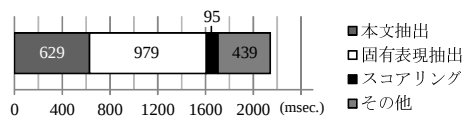


図 3 キーワード抽出の平均処理時間
Fig.3 Mean length of time for terms extraction

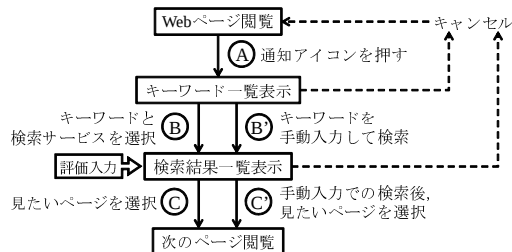


図 4 操作フロー
Fig.4 Operation flow

Mobile) のアドオンとして実装した。なお、検索エンジンには Google の API を用いた。

2.4 キーワード抽出の処理性能

ニュース、ブログなど複数分野の 15URL (ページサイズは平均 89k バイト) に対し、キーワード抽出処理でかかった時間を 5 回ずつ測定した結果の平均を図 3 に示す。キーワード抽出全体で平均 2.14 秒であった。

キーワード抽出処理はユーザが Web ページを閲覧中にバックグラウンドで実行されるため、ユーザにとっては関連情報を検索したいと思った際には、既にキーワード抽出が終了していると考えられる。また実用する際には、長文ページ等で処理時間がかかってしまうのを防ぐため、閾値以上の長さのテキストでは本文抽出や固有表現抽出の途中で処理を打ち切っている。

3. 被験者評価

実際のユーザによる印象を調査するために、IT 系技術者 14 名を対象に利用評価を実施した。

評価の進め方は以下の通りである。また操作フローを図 4 に示す。

- (1) 被験者は約 1 時間、自由に Web ブラウジングを行う。検索をしたいタイミングで通知アイコンを押す、検索支援機能を利用する (A)。検索に利用したいキーワードが表示されている場合は選択して検索し (B)、表示されていない場合は画面上の検索ボックスでクエリを手動で入力して検索する (B')。
- (2) 検索結果の提示は図 1(C) の UI でなく、評価用 UI (図 5) で行われる。提示件数は最大 8 件とし、被験者は検索する毎に以下の評価を入力する。
検索利用目的 検索する前に出てきて欲しい情報の

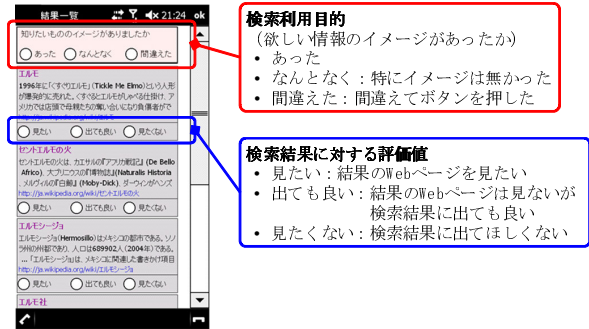


図 5 評価用 UI
Fig.5 UI for evaluation

イメージがあったかどうかを回答する。「あった」(能動的検索)、「なんとなく」(受動的検索)、「間違えた」(間違えてボタンを押した、集計対象外)の 3 種類から選択する。

各検索結果に対する評価値 それぞれの検索結果に対し、被験者の満足度合いに応じて 3 段階で評価する。評価値は以下の通り。

- 3: 見たい (そのページの内容を閲覧したい)
 - 2: 出ても良い (閲覧したいわけではないが、検索結果に含まれていてもよい)
 - 1: 見たくない (結果に表示されてほしくない)
- (3) 検索結果一覧から選択することで、次のページを閲覧する。

通知アイコンを押してから次のページに遷移するまでを 1 セッションとし、全検索結果の中に 3:見たいと評価された結果が含まれる場合、そのセッションの結果は受容されたとする。また、検索支援機能を利用したセッション数 (通知アイコン押下数) を A、キーワード選択による検索セッション数を B、そのうち受容されたセッション数を C とする。関連情報を調べたいと思った際に手軽に満足できる推薦結果を提示できるかどうかを測定するため、以下の評価指標を用いた。

- 提示キーワード利用率 = B/A 。検索したいキーワードが提示された割合を表す。
- 検索結果セッション受容率 = C/B 。キーワード選択後、結果が受容された割合を表す。
- ユーザセッション受容率 = C/A 。検索したいと思った時にキーワード選択による検索で結果が受容された割合を表す。

評価結果を表 1 に示す。最終的に、被験者は検索支援機能を利用した内 62.8% の割合で、選択操作のみの検索によって次の Web ページに到達したこととなる。

評価後、被験者にキーワードを手動入力した理由を確認した。図 6 にその分布を示す。キーワード抽出失

表 1 評価結果 (各数値は被験者毎の平均)

Table 1 Evaluation results (mean values of subjects)

項目	被験者平均
閲覧ページ数	40.0
検索支援機能利用セッション数 (A)	13.9
キーワード選択セッション数 (B, B')	10.2, 1.7
キーワード選択受容セッション数 (C, C')	8.36, 1.64
提示キーワード利用率 (B/A)	73.2%
検索結果セッション受容率 (C/B)	84.7%
ユーザセッション受容率 (C/A)	62.8%

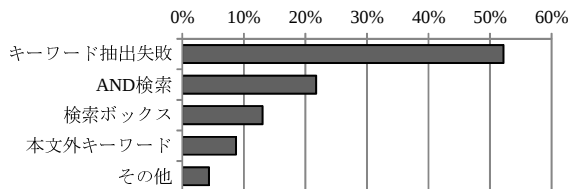


図 6 自分でキーワードを入力した理由

Fig. 6 Reason for manual query input

表 2 検索利用目的毎の検索結果セッション受容率の比較

Table 2 Acceptability for each search intention

項目	能動的検索	受動的検索
キーワード選択セッション数	6.57	3.64
検索結果セッション受容率 (%)	93.0	70.7

敗の他, AND 検索のためや, 本文領域と定義した領域外のキーワードを入力しているケースが見られた。これらは現在の UI 仕様ではサポートしていない。

検索の利用目的別の受容率を表 2 に示す。受動的検索は能動的検索よりは少ないが, 一定数の検索が行われている。一方能動的検索に比べ, 受動的検索では受容率が低い。これは受動的検索時には単純なキーワード検索以上の結果が期待されているためと考えられる。

以上の考察から, 暇つぶしなど受動的検索への潜在ニーズがあることが知見として得られた。課題として, AND 検索などの現在はサポートしていない機能の追加が挙げられる。また受動的検索でも受容できる結果が得られるように, 閲覧領域の考慮やユーザ嗜好への適応なども今後の課題である。

4. 関連研究

提案手法は検索クエリとして有用と考えられるキーワードを抽出するが, 明示的にクエリとして使わない場合も含め, 特定の機能で有用なキーワードを抽出・活用する研究は多数行われている。情報検索の観点では, クエリフリーな情報検索・提示インタフェースが提案されてきた。FIXIT¹⁾ では, クエリフリーでコピー機のマニュアルから修理情報を検索することができる。TV 映像からクエリタームを抽出し, 関連ニュースや Web ページを検索する研究も行われている²⁾⁴⁾。特に

閲覧 Web ページからの関連情報検索では, 補完情報を検索する WeBrowSearch⁷⁾ が提案されている。

単語の意味に応じた検索方式の切り替えでは, 選択した単語の意味分類に基づき検索方法を推薦する手法³⁾ が提案されている。提案手法では意味分類を検索方式の切り替えには用いていないが, 検索方式をユーザが別に選択する UI ではこの方式を組み合わせることも可能である。

5. おわりに

本稿では, 数回の選択操作だけで閲覧中の Web ページの関連情報検索を行うワンタッチ検索支援インタフェースを提案した。また評価を通じ, 実用的な検索精度が得られることを確認するとともに, ユーザニーズに基づく今後の課題を明確にした。

提案手法では Web ページ全体に対して処理が行われるため, 実際には表示されていない部分からもキーワードが抽出される。また, ユーザの閲覧・操作履歴を考慮していないため, 今後はこれらの点も考慮の上, 改良を図る予定である。

参考文献

- 1) Hart, P.E. and Graham, J.: Query-Free Information Retrieval, *IEEE Expert*, Vol.12, No.5, pp.32-37 (1997).
- 2) Henzinger, M., Chang, B.-W., Milch, B. and Brin, S.: Query-free news search, *Proc. of WWW '03*, ACM, pp.1-10 (2003).
- 3) 石谷康人, 鈴木 優, 布目光生: 意味クラス解析と意図推定に基づくインタラクティブな情報検索インタフェース, 情報処理学会論文誌, Vol.48, No.12, pp.3793-3808 (2007).
- 4) Okamoto, M., Kikuchi, M. and Yamasaki, T.: One-button search extracts wider interests: an empirical study with video bookmarking search, *Proc. of SIGIR '08*, ACM, pp.779-780 (2008).
- 5) Rahurkar, M. and Cucerzan, S.: Predicting when Browsing Context is Relevant to Search, *Proc. of SIGIR '08*, ACM, pp.841-842 (2008).
- 6) 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, 服部正典: モバイル機器における閲覧 Web ページからのクエリ抽出を用いた検索支援システム, 電子情報通信学会 Web インテリジェンスとインタラクシオン研究会, Vol.WI2-2010-49, pp.59-64 (2010).
- 7) Yoshida, T., Nakamura, S. and Tanaka, K.: WeBrowSearch: toward web browser with autonomous search, *Proc. of WISE '07*, Springer-Verlag, pp.135-146 (2007).