

TapTapSearch: スマートフォンにおける検索候補語の表示手法の提案

安部 文紀* 寺田 実*

概要. スマートフォンを使った Web ページ閲覧中の単語検索には、範囲指定を利用して単語をコピー & ペーストする方法と、フリック入力などを用いて新たに単語を入力する方法がある。前者は、単語を検索するまでのステップ数が多いうえ、タップ位置の不正確さに起因する誤りが起こりやすい。後者は、入力そのものの遅さが問題である。そこで本研究では、検索単語の近傍をタップすることにより、「タップ位置と単語との距離」と「単語ごとのスコア」の両方を加味して検索候補語をランク付けし、ボタンとして提示する「検索候補語の表示手法」を提案する。提案手法で単語のスコアは「検索対象になりやすさ」を近似するものとし、4 種類考案した。提案手法はブラウザのアドオン TapTapSearch として実装し、有効性を比較検討した。

1 はじめに

スマートフォンを使った Web ページの閲覧中に検索したくなった単語（以下、「検索語」と呼ぶ）を検索する方法は 2 通りある。ひとつは検索語をコピー & ペーストする方法で、もうひとつはフリック入力でブラウザの検索窓に検索語を直接入力する方法である。前者は指をスライドさせて行う範囲指定を利用するため、検索語を検索するまでのステップ数が多いうえ、タップの不正確さに起因する範囲指定の誤りが起こりやすい問題がある。後者は入力そのものの遅さが問題であり、大塚らが実証している [1]。

上述の問題を解決するため、鈴木ら [2] はキャレットを文中に配置し、文全体を指で動かしてキャレットの相対位置を変化させることで文字列範囲を指定する関節操作手法を提案した。これによってタップ位置の不正確さを回避したが、キャレット操作が 2 度（キャレットの初期位置調整と文字列範囲選択）必要であるため、文字列をコピーするまでのステップ数が多い。三浦ら [3] は通常の文字列範囲指定でキャレットを文字単位でなく形態素単位で動かす手法により作業時間の短縮を図ったが、キャレットの設置と操作はタップ位置の不正確さによる影響を受ける。

そこで本研究では、検索語をタップ操作のみで、かつ検索語の正確な位置でタップしなくとも選択できることが望ましいと考え、「タップ位置と単語との距離」と「単語ごとのスコア」の両方を加味して検索語の候補（以下、「検索候補語」）をランク付けし、ボタンとして提示する「検索候補語の表示手法」を提案する。タップ位置の不正確さは意味的な処理によって軽減できるという仮定に基づき、提案手法で単語のスコアは「検索対象になりやすさ」を近似するものとし、4 種類考案した。提案手法はスマートフォンと

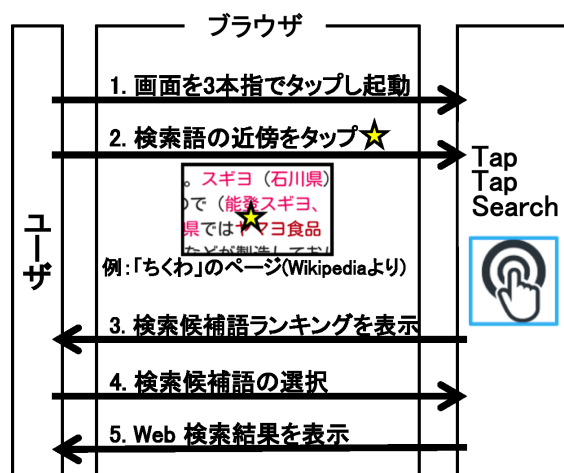


図 1. TapTapSearch の概要図

PC のブラウザで動作するアドオン TapTapSearch として実装し、「タップ位置」に「単語のスコア」を加味することの有効性を比較実験により検討した。

2 TapTapSearch

TapTapSearch の概要図を図 1 に示す。TapTapSearch は通常閲覧モードと検索モードがあり、(1)3 本指で画面をタップして検索モードを起動する。(2) 検索語の近傍をタップすると、(3) 検索対象になりやすい検索語（検索候補語）を表示する。具体的にはタップ位置から近く、「検索対象になりやすさ」が大きい単語ほどスコアを良くし、検索候補語をスコアの良い順にボタンとして検索候補語ウィンドウに 10 個表示する（図 2）。(4) 検索候補語から検索語を選択し、虫眼鏡アイコンを押下して Web 検索する。(5) 検索語の Web 検索結果上位 10 件を表示する。

TapTapSearch が検索モードのときは画面をタップすると検索候補語の表示処理に入るため、Web

Copyright is held by the author(s).

* 電気通信大学 情報理工学研究所 情報・ネットワーク工学専攻

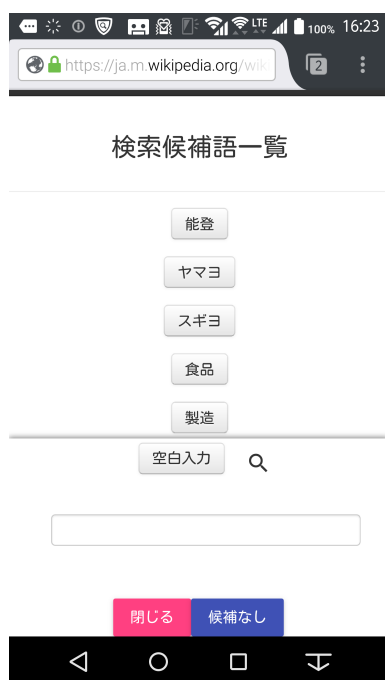


図 2. 検索候補語ウィンドウ (スクロール可能)

ページのハイパーリンクをたどる場合は3本指で画面を再びタップして通常閲覧モードにする。

3 実装：検索候補語の決定 (スコアリング)

- 距離スコア: タップ位置の座標から単語の中心座標までの画面上の2次元距離
- 単語スコア: 単語の検索対象になりやすさ
- 総合スコア: 距離スコアを単語スコアで除算したスコア

本研究における単語スコアは「検索対象になりやすさ」であって「ページ中での重要語」とは異なる。例えば“ちくわ”をWeb検索して辿り着くページはちくわが重要語であり、ちくわを主題にした内容のはずである。この時ちくわのページを閲覧中に検索したい単語はちくわとは別の単語のはずである。この検索対象になりやすさを以下に示す4つの方式で近似した。また、総合スコアが小さい単語ほど検索対象になりやすく、検索候補語の順位が高いとする。

方式 1:

$$\text{単語スコア} = TF \times \text{WebIDF} \quad (1)$$

方式 2:

$$\text{単語スコア} = \text{WebIDF} \quad (2)$$

方式 3:

$$\text{単語スコア} = \frac{1}{TF} \quad (3)$$

方式 4:

$$\text{単語スコア} = \frac{1}{PMI \times \text{Discounting Factor}} \quad (4)$$

方式 1 は閲覧ページに頻出する単語で、かつ多くの Web ページに存在しない単語は検索対象になりやすいとする。方式 2 は多くの Web ページに存在しない単語は検索対象になりやすいとする。方式 3 は閲覧ページに頻出しない単語は検索対象になりやすいとする。方式 4 は閲覧ページのタイトルとの関連度が低い単語、つまり Web ページの主題から遠い単語は検索対象になりやすいとする。TapTapSearch には、比較実験で「検索対象になりやすさ」を最も近似した方式 2 を採用することとした。

4 まとめ

本研究ではスマートフォンを使った Web ページの閲覧中に検索語をタップ操作のみで、かつ単語の正確な位置でタップしなくとも選択できる「検索候補語の表示手法」を提案し、スマートフォンと PC で使用可能な Web ブラウザのアドオン「TapTapSearch」としてそれを実装した。本研究のアイデアはタップ位置と検索語との距離が離れていても検索語を選択できることである。そのため、提案手法は検索語の近傍をタップすることによって「タップ位置と単語との距離」と「単語ごとのスコア」の両方を加味して検索候補語をランク付けし、ボタンとして提示する。

本研究の議論は Web ページにおける「重要な単語」と、ユーザが検索したい単語すなわち「検索されやすい単語」は異なるのではないかという点である。そこで、提案手法で単語スコアは「検索対象になりやすさ」を近似するものとして4種類考案した。

実験では提案手法の有効性を2つの視点から検討した。ひとつは、単語スコアを用いることで検索語を離れた位置からタップしても選択できること。もうひとつは、Web ページ中で重要な単語は必ずしも検索対象になりやすい単語ではないことである。結果は、単語スコアを用いることで検索語を離れた位置からでも選択できることを確認し、検索対象になりやすさを近似するには単純な Web ページ中の単語重要度よりも有効なものがあることを確認できた。

以上より、提案手法の「タップ位置」に「単語のスコア」を加味することの有効性を確認できた。また、TapTapSearch の単語スコアには検索対象になりやすさを最も近似した方式 2 を用いることとした。

参考文献

- [1] 大塚正雄 他. “モバイル機器の文字入力の使用やすさ: モバイル機器の文字入力速度の比較検討”. 広島国際大学医療経営学論叢, 2012.
- [2] 鈴木健司 他. “Fix and Slide: 文字列全体を操作することによるキャレットの操作手法”. WISS, 2015.
- [3] 三浦元喜 他. “タブレット端末における日本語形態素解析を利用した文書範囲選択手法”. 情報処理学会論文誌, Vol.57 No.4, pp. 9991–9999, 2016.