

少数キーでの入力可能な携帯端末向け Web 検索手法の有効性について Effectiveness of Web Search Method Using 12 Keys on Mobile Terminal

松原 雅文[†]
Masafumi Matsuhara

1. まえがき

近年、端末自体の性能向上や技術進歩、さらには、通信インフラの整備に伴い、スマートフォンに代表されるように、インターネットに接続可能な小型携帯端末が増加している。そのため、電子メールや SNS の利用もさることながら、このような携帯端末上で、Web 検索を行う機会と必要性が増大している。

これに対して、我々は、携帯端末向け Web 検索手法を提案している [1]。本提案手法では、基本的に少数のキーしか装備していない携帯電話などの小型端末において、検索クエリーの入力処理も含め、全体として高速な Web 検索を実現することを目指している。クエリー文字列の入力を曖昧さを残した状態で完了することにより、画面注視を抑制し、かつ、迅速な入力を可能とする手法である。この入力の曖昧さは検索の過程で自動的に解消する。評価実験の結果より、携帯端末上での通常の Web 検索手法と比較して、打鍵数の減少が確認できたことから、その有効性が示された。

しかしながら、その検索精度は 44[%] 程度と低い値であった。これは、意図したキーワードが辞書に未登録であったり、入力の曖昧さに起因するものであった。本手法では、使用者に合わせた辞書を作成するため、Web 検索結果において使用者が選択した Web ページから、随時、単語を獲得している。よって、辞書に未登録であった単語は、その時点で学習されることとなり、次回以降のキーワード候補となることが可能である。ここで用いられたデータは、著者自身の 40 日分の Web 検索履歴のうち、180 件程度であった。獲得された単語が、次回以降の検索において、どの程度有効に作用するかを確認するためには、より長期間のデータを用いた調査が必要であると考えられる。

そこで、本稿では、提案手法の概要を説明し、より長期間の Web 検索履歴データを用いて評価を行った結果から、本手法の有効性について述べる。

2. クエリーの入力

Web 検索を行うためには、通常、複数のキーワードをクエリーとして与える。このクエリー文字列は、ふつう、キーボードから入力される。よって、この入力処理自体も、小型端末上で行われる必要がある。スマートフォン等の小型端末上での文字入力方式としては、フリック入力やソフトウェアキーボードによるものが考えられるが、その操作に戸惑ったり、キーの小ささに起因して入力誤りが増える、といった問題点がある。

そこで、いわゆるフィーチャーフォン等の携帯電話にも搭載されている数字キーを用いた入力を考える。文字循環指定方式（マルチタップ方式）と呼ばれる入力方式においては、1 文字の入力に複数回の打鍵が必要となり、迅速な入力は難しい。そこで、文字情報縮退方式（シングルタップ方式）[2] と呼ばれる入力方式の適用を考える。この方式において、ユーザは、意図したキーワードの各文字に対応した数字を 1 文字につき、1 打で入力可能である。よって、クエリー文字列の入力を迅速に行うことができる。その反面、この入力方式において入力された数字列は、曖昧さを含むこととなる。

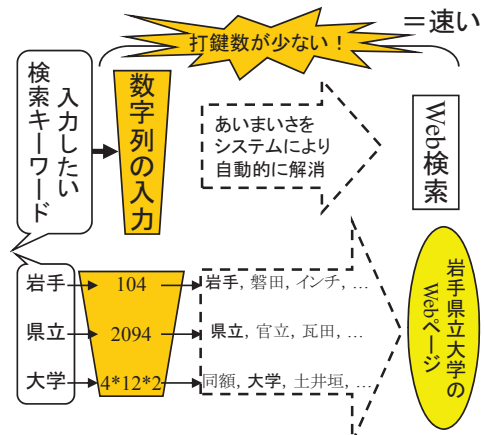


図 1: 提案手法

シングルタップ方式を用いた日本語入力手法が提案されている [3], [4]。よって、このような手法を用いることにより、入力数字列の曖昧さを排除し、これを漢字かな文字列へと変換することが可能である。しかしながら、この際には、変換、確定処理に伴う打鍵や画面注視が必須となる。また、高精度の変換を行うためには、文脈情報が必要となるが、Web 検索におけるクエリーは、基本的には、短いキーワードから構成される。よって、この文脈情報を利用するのが難しいことから、変換精度の低下も懸念される。さらに、Web 検索の目的は、意図した Web ページを取得することであり、必ずしも、漢字かな文字列への変換処理は必要ないものと考えられる。

そこで、本手法においては、クエリーを人手によるコストをかけて曖昧さの低い漢字かな文字列に変換することは行わず、入力された曖昧さを含む数字列から、直接、Web 検索を行うものとしている。これにより、検索クエリーの入力処理も含め、全体として高速な Web 検索を実現することを目指している。

3. 携帯端末向け Web 検索手法

本提案手法の概念図を図 1 に示す。Web 検索時のクエリー入力を小型携帯端末上の少数キーのみを用いて、かつ少ない打鍵数で完了させるため、本手法においては、シングルタップ方式を採用し、キーワードの各文字に対応した数字を 1 文字につき 1 打で入力する。

この各数字キーへの文字の割り当てを表 1 に示す。表 1 のとおり、本手法において、各数字キーには、かな文字だけでなく、アルファベットや数字自体も同時に割り当てている。この割り当ては、一般的な携帯電話における各数字キーへの文字の割り当てと同様であることから、一般ユーザにとっても利用しやすいものと考えられる。さらに、本手法においては、これらの文字種を切り替える必要はなく、すべて 1 文字 1 打で入力が完了することから、迅速な入力が可能である。

図 1 においては、意図したキーワード「岩手」に対応した数字列を入力している。従来手法であるマルチタップ方式により同様の文字列を入力する場合を考えると、かな文字列「いわて」に対応した数字列として「1104444」を入力することとなり、多くの打鍵数が必要である。さらに、これを漢字に変換し、意図した文

[†]岩手県立大学, Iwate Prefectural University

表 1: 各数字キーへの文字の割り当て

1:あいうえお 1	2:かきくけこ ABC 2	3:さしすせそ DEF 3
4:たちつてとっ GHI 4	5:なにぬねの JKL 5	6:はひふへほ MNO 6
7:まみむめも PQRS 7	8:やゆよやゆよ TUV 8	9:らりるれる WXYZ 9
*:(半)濁点, 記号	0:わをん 0	#:空白 (単語区切り)

字列を確定する必要があることから、より多くの打鍵と画面注視が必須となる。

これに対して、本手法においては、キーワード「岩手」を入力するため、かな文字列「いわて」の各文字に対応した数字を入力しており、数字列「104」が入力されている。さらに、文字列の変換、確定処理が不要なことから、あまり画面を注視する必要もない。

一方で、本手法において入力された数字列は、多くの曖昧さを持つこととなる。図 1 では、入力数字列「104」に対応するキーワードは、「岩手」の他にも「磐田」「インチ」など多数が競合している。これをユーザにより選択してもらい、検索前にキーワードを確定する方法も考えられる。しかしながら、本手法においては、この曖昧さは、検索の過程で、システムにより自動的に解消するものとする。具体的には、複数キーワード「岩手」「県立」「大学」が同一 Web ページ上に同時に出現する機会が多いのに対して、「磐田」「官立」「同額」などが同一 Web ページ上に同時に出現する機会は比較的少ないものと考えられる。

このような共起情報等を利用することにより、本手法においては、数字列の曖昧さを解消し、ユーザが意図した適切な Web ページを迅速に提示することを目指している。

4. 評価

本手法の有効性を確認するため、Web 検索履歴データを用いて、これに含まれる単語の評価を行った。

4.1. 評価データ

評価データとして、著者自身の Google[†] における実際の Web 検索履歴のうち、2014 年の 1 年分すべてを用いた。総クエリー数は 2,690、これに含まれる延べ単語数は 4,600、異なり単語数は 2,300 であった。

4.2. 評価方法

評価データについて、各単語を数字列として表し、その出現頻度を求めた。

本手法においては、1 度目の出現時に単語が獲得されるので、次回以降の検索では、これをキーワード候補として出力することが可能である。よって、2 回以上出現している単語は、基本的には、Web 検索時のキーワード候補となり得る。しかしながら、本手法においては、入力数字列の曖昧さから、キーワード候補が競合する、すなわち、2 種類以上の異なる単語として認識される可能性がある。キーワード候補が競合すると、適切な Web 検索を行えない場合がある。

そこで、2 回以上出現した数字列について、どの程度、競合が発生しているかを評価した。

4.3. 評価結果および考察

Web 検索履歴を評価した結果を表 2 に示す。異なり数字列数と、このうち 2 回以上出現した数字列の個数とその割合、および、この 2 回以上出現した数字列のうち、複数の異なる単語として解釈されるもの、すなわち、競合が発生している数字列の個数とその割合を表している。

全体として、7 割程度は、1 回のみしか出現していないことから、現状では対応が困難であり、汎用的な辞書を利用する等、検討が必要である。残りの 3 割程度に関しては、2 回以上出現していることから、当該単語

表 2: 評価結果

数字列数	2 回以上出現した数字列数 (%)	競合が発生した数字列数 (%)
2,156	713 (33.1)	123 (17.3)

を 1 度目の出現時に獲得しておけば、基本的には、これを、次回以降の Web 検索時にキーワード候補として出力することが可能である。

しかしながら、数字列の持つ曖昧さから、17[%] 程度の数字列において競合が発生していた。この競合が発生している数字列は、複数の異なる単語として認識されることから、Web 検索時に意図した結果が得られない可能性がある。しかしながら、この競合が発生した数字列の中には、その差異がアルファベットにおける大文字と小文字の違いだけであるものが 59 個存在した。これらは、Web 検索時には、基本的に同一の結果となることから、競合が発生していないと見なすことができる。よって、これを除いた 64 個のみに競合が発生していることとなり、この割合は 1 割未満となる。

このように、現状の手法において対応している出現頻度が 2 以上の単語について、これが曖昧さを持つ数字列として入力された場合であっても、その 9 割以上について、適切な Web 検索結果を出力できることが示唆された。

5. おわりに

本稿では、少数キーでの入力が可能な携帯端末向け Web 検索手法を説明し、長期間の Web 検索履歴データを用いて行った評価結果から、本手法の有効性について述べた。

迅速な入力のため、本手法において入力されるクエリーは、曖昧な数字列となっている。この曖昧さは、検索時に解消するものとしているが、これがどの程度可能かを、実際の Web 検索履歴を用いて評価した。結果として、2 回以上出現している単語について、その 9 割以上は適切な Web 検索結果を出力できることが示唆され、本手法の有効性が確認された。

今回の評価データは、1 人分の Web 検索履歴であったので、今後は、複数のユーザにおける Web 検索履歴を収集し、その評価を行っていく予定である。また、1 回のみしか出現していない単語についてもキーワード候補として出力すべく、汎用的な辞書の利用等についても検討していく必要がある。

謝辞

本研究の一部は JSPS 科研費 15K00155 の助成を受けたものである。

参考文献

- [1] 松原 雅文, “帯端末向け Web 検索手法の有効性について,” 情処研報 (2005-NL-170), pp.39-44, November 2005.
- [2] 東田 正信, 奥 雅博, 村上 仁一, “PB 電話機を用いた自動電話番号案内システムの開発と評価,” 信学論 D, Vol.J96-D, No.10, pp.2249-2261, October 2013.
- [3] 松原 雅文, 荒木 健治, 柄内 佳雄, 柄内 香次, “文字情報縮減方式を用いた帰納的学習によるベタ書き文の数字漢字変換手法の有効性について,” 信学論 (D-II), Vol.J83-D-II, No.2, pp.690-702, February 2000.
- [4] 田中久美子, 犬塚祐介, 武市正人, “携帯電話における日本語入力—子音だけで日本語が入力できるか,” 情処学論, Vol.43, No. 10, pp.3087-3096, October 2002.

[†]<http://www.google.com>