

周辺語を活用したクリック型 Web 検索システムの提案と評価

服部 元^{†1} 原 隆浩^{†2} 滝嶋 康弘^{†1}
菅谷 史昭^{†1} 西尾 章治郎^{†2}

近年、携帯電話やゲーム機、カーナビゲーションシステム、テレビなど、Web ブラウザを搭載している PC 以外の機器が増加している。しかしながら、それらの機器は一般にキーボードを搭載せず、数字ボタンや十字キーなどの限られた入力インタフェース（以下、リモコン型 UI と呼ぶ）のみを備えているものが多い。そのため、Web ページの閲覧は容易に行えるものの、Web ページの検索を行う場面では、試行錯誤しながら検索語入力を行い検索結果を絞り込むという PC と同等の検索手法は困難であることから、リモコン型 UI 端末向けの使いやすい Web 検索方法が求められている。そこで本論文では、閲覧中の Web ページに表示されている語の中からユーザが検索したい語（注目語）を指定すると、システムが自動的に最適なページを提示するクリック型検索手法を提案する。本手法は、検索に重要な語を注目語の周辺から抽出することを特徴としており、品詞や意味、表示方法などに基づく語単独の重要度と、注目語との距離や共起の度合いに基づく重要度の 2 つの指標からなる周辺語重要度モデルを利用して、語の重要度を算出している。本手法を実装して検索結果に対する主観評価実験を行った結果、注目語のみで行った検索において評価が低かった検索結果のうち、最大で 63.6% のケースにおいて満足度が向上することを示した。また、2 クリック以内の検索結果に対しては、従来方式のみの場合と比較して、満足するケースが最大で 63.3% から 76.0% に向上することを示した。

Proposal and Evaluation of a Click-based Web Search System Using Surrounding Words

GEN HATTORI,^{†1} TAKAHIRO HARA,^{†2}
YASUHIRO TAKISHIMA,^{†1} FUMIAKI SUGAYA^{†1}
and SHOJIRO NISHIO^{†2}

The Internet today provides an ubiquitous environment for viewing Web pages. Now we can view them on many kinds of terminals such as mobile

phones, video game machines, car navigation systems, and television sets, besides PCs. On the other hand, most of them have limited input devices which consist of 9 number buttons and 4 direction buttons, which we call “Remote controller type input devices.” This limitation causes a difficulty to the Internet search task that we have to input correct key words by trying and erring until we find the best Web page. Therefore, we need a search system to make the task easier for Remote controller type input devices. This paper proposes a new search system, click based Web search system, which automatically searches related Web pages with a core word and surrounding words when a user clicks a specified word (the core word). The most important technique used in this system is a method to calculate the importance of each surrounding word based on a surrounding word weighting model. This model consists of the importance based on their parts of speech, meanings and expression, and the importance based on the distance and cooccurrence between a core word and the word. We performed subjective evaluation and showed that the proposed method have achieved a satisfaction rate of 63.6%, which is much higher than the results produced by other common methods on low evaluated cases. When applying to the second click the proposed method helps to increase satisfaction rate from 63.3% when only using common method to 76.0% of the final result.

1. はじめに

近年、携帯電話やゲーム機、カーナビゲーションシステム、テレビなどの情報家電端末に Web ブラウザが搭載されるようになり、ユーザはそれらの端末を利用してインターネットの情報を見られるようになった。このように、PC 以外の Web 閲覧環境が増加しており、閲覧する場所を限定されることなく、また PC の操作に詳しい人でなくても、誰でも気軽にインターネットの情報にアクセスできる環境が整いつつある。

しかしながら、これらの端末で Web 閲覧を行うためには、ユーザは数字キーや方向キーなどの限られた入力インタフェース（以降、リモコン型 UI と呼ぶ）を利用することになる。このため、キーボードやマウスを有する PC と同様の操作性を実現するのは難しい。ここで、マウスについては、一部の携帯電話端末にマウスポインタが搭載されていることや、携帯電話上でマウスを動作させるソフトウェア¹⁾が開発されていること、さらに一部の TV ゲーム機のブラウザではポインタが利用できるなど、ポインタの搭載は進んでいる。一方、

^{†1} 株式会社 KDDI 研究所
KDDI R&D Laboratories Inc.

^{†2} 大阪大学情報科学研究科
Graduate School of Information Science and Technology, Osaka University

キーボードについては、個々のキーを指で押す必要があることから、ある程度の設置面積が必要であるといった物理的な制約がある。そのため、一部の高性能な携帯電話端末や PC 一体型のテレビ端末に搭載されているのみであり、ほとんどの端末はキーボードを搭載していない。画面上に表示されたソフトウェアキーボードをポインタでクリックしながら文字入力する手段はあるが、物理的なキーボードほどの快適な文字入力はできない。

一方、多くの Web 検索サービスは日々進化しており、少ない検索語数でもまったく見当違いの Web ページが提示されることは少なくなった。しかしながら、基本的には多くの人が望むと推定した Web ページを優先的に提示するアルゴリズムであるため、ユーザが稀な情報を探している場合は、複数の適切な検索語を選び、試行錯誤をしながら目的の情報を探し当てる作業を行うことになる。さらに、OneStat.com 社の調査結果によると²⁾、Web 検索時に入力する検索語数で利用頻度が最も多いのは 2 語であり、また 10 語までの頻度分布で平均すると約 3 語となる。よって、望みの情報を得るためには多くの文字を入力する必要があるといえる。

以上のことから、リモコン型 UI のみの情報家電端末で、一般的な Web 検索操作を満足に行うことは非常に困難であるといえる。よって、リモコン型 UI を利用した Web 検索を容易に行えるようにするために、検索語入力の手間を削減する新たな Web 検索手法を検討する必要がある。ここで、ユーザの検索行動を、任意の語で検索する場合と、表示中の情報に関連する語で検索する場合を組み合わせたと仮定する。このとき、前者についてはユーザの頭に浮かんだ検索語を入力する必要があるため、文字入力操作は避けられない。一方、後者については、検索の足がかりとなる語が必ず Web ページ中に表示されることとなり、ユーザの検索語入力操作を削減することができる可能性が高い。

そこで本論文では、ユーザが Web ページ上で注目した語（以降、注目語と呼ぶ）をポインタでクリック操作をすると、その前後にある語（以降、周辺語と呼ぶ）の重要度に基づき検索語を動的に生成して検索するクリック型検索手法を提案する。本手法は、周辺語の重要度を、品詞や意味、表示方法などに基づく語の単独での重要度と、注目語に対する関連の強さに基づく重要度の 2 つの指標を用いて計算する。

2. クリック型 Web 検索手法の機能要件と目標

本論文では、ポインタが利用できることを前提とし、ユーザがリモコン型 UI を利用して、容易に検索操作ができるクリック型 Web 検索システムの実現を目標とする。ここで、リモコン型 UI の操作性については、たとえば赤外線や加速度センサを利用したポインティング



図 1 リモコン型 UI を利用した検索語指定操作の例

Fig. 1 Example of a remote controller type input device for selecting a keyword on a Web page.

デバイスの場合は、画面上の任意の位置に素早くポインタを移動することができるという利点がある。しかしながら、ドラッグ操作で必要なポインタの位置を正確に水平または垂直方向に移動させることや、一定位置にポインタを維持しながらクリック操作をすることが難しいという欠点がある。また、携帯電話に適用されている方向キーを利用したポインティングデバイスの場合は、ポインタの位置を正確に移動・固定することができるという利点があるが、任意の位置に素早く移動させることが難しいという欠点がある。

以上の検討結果から、リモコン型 UI 向けのクリック型 Web 検索システムは、検索に必要なクリック数を最小限にすることが機能要件となる。本論文では、ユーザが指定した語に対し、その語を修飾できる「重要語」を周辺から判別して、適切な検索語を自動生成することで機能要件を満足する方針とする。想定するクリック型 Web 検索システムの利用シーンを図 1 に示し、ユーザの検索操作手順を以下の (1)~(3) に示す。本論文ではまず、最小のクリック数である 1 回のクリックで検索する方式について提案し、基本的な性能評価を行う。次に、2 クリック以上を用いて検索の満足度を上げる方法について、実装方法の検討と合わせて議論する。

- (1) ある Web ページが画面上に表示されているとする。この Web ページはユーザのブックマークやポータル設定、あるいはそれらからリンクをたどるなどの手順を行い、表示されているものとする。
- (2) 画面上に表示されているテキストの中に検索したい語があった場合に、ユーザはポインタをその語に重ねるように操作し、次にボタンをクリックする。これにより、ユーザはシステムに対し検索語を指定する。
- (3) リスト形式やサムネイル表示、あるいは最適と推定した Web ページを直接表示する

など、何らかの形式で検索結果が画面上に表示される。ユーザは提示された検索結果の中から望みの Web ページを探す。

3. 関連研究

重要語の推定手法ならびにキーボードを多用しない検索手法について、関連研究とその課題を述べる。

重要語の推定手法として、語の共起情報を利用して重要な語を文書から抽出する手法³⁾や、多義語に代表される検索語の曖昧性を解消するために、検索結果から重要語を抽出する手法⁴⁾がある。前者は、コーパスや単純なルールを利用する重要語推定方式^{5)~7)}では Web 文書の多様性には対応できないことを課題として、Web 文書単独での重要語推定方式を提案している。また後者は、一部の語とだけ共起する語を重要語とする手法である。これらの手法は単独あるいは限られた Web ページ集合から重要度を推定する点で、我々の着目している課題と類似しており、語の重要度の 1 つの要素として活用することができる。しかしながら、これらが抽出しているのは 1 つの Web 文書あるいは限られた Web 文書集合内における重要度のみであり、ユーザが着目した語に対する重要度に関する検討がないという課題がある。

ペンインタフェースと検索目的指定による検索操作補助手法⁸⁾は、タッチパネルとペンによる入力インタフェースを利用して、ユーザがペンのドラッグ操作で検索語を選択すると、システムがその周辺にある語の種別（企業、施設など）に応じた検索意図メニュー（地図が見たい、ニュースリリースが見たいなど）をユーザに提示する。この方式は、ユーザの検索意図を確実に絞り込むことができるという利点はあるが、リモコン型 UI では正確に文字列をなぞるのが難しいことや、検索に至るまでの入力の手数が多いこと、さらに語の種別を判別するための辞書および検索意図知識をあらかじめ構築しておき、定期的に更新するなどの対応が必要であることが課題である。

また、検索結果を分類しながら絞り込んでいく検索手法^{9),10)}は、検索結果を内容に基づきいくつかのグループに分類し、ユーザがそのうちの 1 つを選択することで検索目的を伝える方式である。グループを選択するとさらにその中の Web ページを詳細なグループに分類していき、このインタラクション操作を繰り返すことで絞り込みを進めていく。しかしながら、検索目的を段階的に指定していくことになるため、検索結果が絞り込まれるまでにはインタラクションが何度も必要となり、検索結果の数によっては入力の手数が大幅に増える課題がある。

4. 提案手法

2 章で述べた機能要件を満たす、リモコン型 UI 向けのクリック型検索手法を提案する。クリック数を最小限にするため、1 回のクリックで最適と思われる 1 つの Web ページを提示する提示型ユーザインタフェース¹¹⁾とし、それ以外の検索結果を参照する場合は、クリック操作のみで順次 Web ページを提示する方針とする。また、クリック数を削減するため、検索結果を後で絞り込む方針^{4),9)}ではなく、検索語を自動的に追加して絞り込む方針¹¹⁾とする。ここで、周辺語は注目語を修飾する重要な語であると仮定し、ユーザが指定した注目語の周辺に表示されている周辺語から重要度の高い語を選択して検索語を自動生成する。

4.1 節および 4.2 節で、クリック型検索手法の処理フローと周辺語を選択する指標である重要度モデルについてそれぞれ述べる。

4.1 クリック型検索手法

提案方式の処理フローを図 2 に示す。ここで (Step3) で抽出する周辺語数は予備実験を行い決定する。また、予備実験については 5.1 節で詳細を述べる。

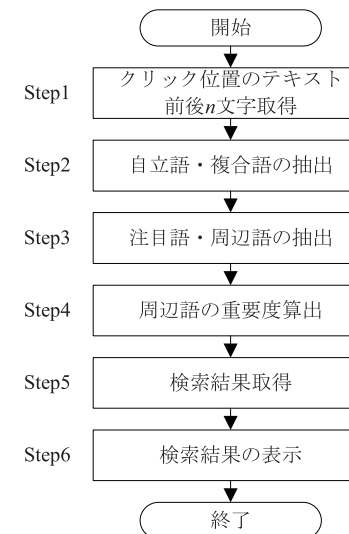


図 2 システムフロー
Fig. 2 System flow.

図 2 の各ステップについて詳細を以下に述べる．

- (Step1) クリック位置の前後にある文字列の取得
 ユーザがクリックした位置に最も近い文字を中心として、前後 n 文字を切り出す．
- (Step2) 自立語・複合語の抽出
 形態素解析ツール¹²⁾を用いて名詞に属する語を抽出する．また、複合語の生成ルールとして、連続する名詞は複合語とし、特定の接尾辞の前では区切ることとする．ここで特定の接尾辞とは、「サ変接続、一般、形容動詞語幹、助数詞、助動詞語幹、人名、地域、特殊、副詞可能」である．また、切り出した文字列の両端にある語は、不完全である可能性が高いため、ここで除外する．
- (Step3) 注目語・周辺語の抽出
 (Step2) で抽出した語の中から、クリック位置に表示されている語を注目語とし、それ以外の語を周辺語とする．
- (Step4) 周辺語の重要度算出
 抽出した周辺語について、注目語と組み合わせて検索することで絞り込みができるような語を選択する．選択方法は、4.2 節で詳細を述べる．
- (Step5) 検索結果取得
 注目語と、(Step4) で選択した周辺語を組み合わせて、一般の検索サイトが提供している検索 API¹³⁾を利用して検索結果を取得する．
- (Step6) 検索結果の表示
 検索結果の Web ページを上位から順にユーザの端末に表示する．

4.2 周辺語の重要度モデル

ある周辺語について、注目語に組み合わせて検索語とすることで、検索結果の絞り込みができる可能性の度合いを“周辺語重要度”と定義する．周辺語は注目語の意味を補完できる関連性の高い語である必要がある．そこで本論文では、

- (1) 語自体の重要度：その語が単独で重要な意味を持つこと、
 - (2) 注目語に対する重要度：注目語との関連性が高いこと、
- の 2 つの指標で周辺語を評価し、その加算値が高い値を持つ場合に周辺語重要度が高いと判定するモデルを提案する．提案モデルを式 (1) に示す．

式 (1) において、注目語 w_a に対する周辺語 w の周辺語重要度 $E(w, w_a)$ は、第 1 項の w における語自体の重要度 $E_o(w)$ と、第 2 項の w_a に対する w の重要度 $E_r(w, w_a)$ の重み和とする．また、 α は第 1 項と第 2 項の加算比率を表す重み係数である．

表 1 周辺語重要度の要素一覧
 Table 1 Elements included in $E(w, w_a)$.

| 重要度種別 | 要素 |
|----------------------------|---------------------------------|
| $E_o(w)$: 語自体の重要度 | $F_m(w)$: 意味 (地名, 人名など) |
| | $F_v(w)$: 表示 (括弧囲み, 強調) |
| $E_r(w, w_a)$: 注目語に対する重要度 | $F_c(w, w_a)$: 共起 (注目語と共起する頻度) |
| | $F_d(w, w_a)$: 距離 (注目語との間の文字数) |

w_a : 注目語 w : 周辺語

$$E(w, w_a) = \underbrace{\alpha \cdot E_o(w)}_{\text{第 1 項}} + \underbrace{(1 - \alpha) \cdot E_r(w, w_a)}_{\text{第 2 項}} \quad (1)$$

4.3 重要度の算出

$E_o(w)$ および $E_r(w, w_a)$ を測定するための、具体的な要素の例を表 1 に示す．ここでは、ニュースや史実などの文書に含まれる注目語を対象とし、周辺語として関連する人物や事象の場所などの固有名詞が利用可能である状況を想定する． $E_o(w)$ の要素として、地名や人名などの文書の内容を限定しやすい固有名詞であるか否かのフラグ $F_m(w)$ 、および強調表現などの Web ページの著者が意図的に重要であると表現した語であるか否かのフラグ $F_v(w)$ の 2 つとする．また、 $E_r(w, w_a)$ の要素として、注目語との共起の度合い $F_c(w, w_a)$ 、および注目語との間の文字数に基づき算出する距離 $F_d(w, w_a)$ の 2 つとする．

以下に、表 1 の要素を持つ場合の式 (1) の第 1 項および第 2 項の算出例を述べる．
 (1) $E_o(w)$ の算出

$E_o(w)$ は、 $F_m(w)$ 、 $F_v(w)$ を含む式 (2) に基づき算出する． $p(w)$ は w の統計的な重要度を表す関数とし、ここでは語の一般的な重要度の指標である idf ¹⁴⁾ を利用する．

$$E_o(w) = (F_m(w) + F_v(w) + 1) \cdot p(w) \quad (2)$$

式 (2) の $F_m(w)$ 、 $F_v(w)$ 、 $p(w)$ は、それぞれ式 (3)、式 (4)、式 (5) で定義する．ここで、 D_M は形態素解析辞書が持つ名詞種別の辞書を表し、 D_B は括弧 (「」, () , “ ”【】 , 『 』 , など) および HTML タグによる強調表現 (, など) を含む、言語にほとんど依存しない小規模な辞書を表す． $df(w)$ は w が出現するドキュメント数を表し、 N はそのドキュメント集合に含まれる全ドキュメント数を表す．

$$F_m(w) = \begin{cases} 1 & (w \in D_M) \\ 0 & (w \notin D_M) \end{cases} \quad (3)$$

$$F_v(w) = \begin{cases} 1 & (w \in D_B) \\ 0 & (w \notin D_B) \end{cases} \quad (4)$$

$$p(w) = \log_2 \frac{N}{df(w)} \quad (5)$$

(2) $E_r(w, w_a)$ の算出

共起語は注目語からの距離が離れるほど関連する度合いが小さくなると仮定する。 $E_r(w, w_a)$ は、 $F_c(w, w_a)$ を $F_d(w, w_a)$ で割った値とし、式 (6) に従い算出する。

$$E_r(w, w_a) = \frac{F_c(w, w_a)}{F_d(w, w_a)} \quad (6)$$

式 (6) の $F_c(w, w_a)$ と $F_d(w, w_a)$ は、それぞれ式 (7) と式 (8) で定義する。 $freq(w)$ は w の出現数を、 Doc は表示中の Web ページ中のテキスト全体を表す。また $dist(w, w_a)$ は、 w と w_a の間にある文字数に基づく表示上の距離を表す。 k は w に対して共起する w_a の候補を選択する範囲とする。

$$F_c(w, w_a) = \frac{freq(w|dist(w, w_a) < k)}{freq(w|w \in Doc)} \quad (7)$$

$$F_d(w, w_a) = dist(w, w_a) \quad (8)$$

5. 評価

提案方式の有効性を評価するため、実際の Web ページを利用した評価実験を行った。4.3 節で述べた要素に基づく提案方式を実装したシステムを構築し、従来方式と提案方式それぞれについて 3 人の被験者による主観評価実験を行った。なお、本章での評価の目的は周辺語を利用した Web ページ推薦の精度を測ることであることから、リモコン型 UI の形状や機能、方式などの評価への影響を避けるため、ポインティングデバイスとして一般的なマウスを利用した。また、クリック位置がずれてしまい、意図しない語を注目語としてしまうなどのクリック失敗の度合いについては、リモコン型 UI の種類に依存するため、本評価では検証しない。

まず、評価用の注目語を収集するための実際の Web ページとして、表 2 に示す 5 つの力

表 2 評価対象のカテゴリ
Table 2 Categories for evaluation.

| カテゴリ名 | 概要 |
|-------|---------------------------|
| IT 系 | IT 関連の情報を提供している Web ページ |
| メディア系 | テレビ局などが提供している Web ページ |
| 医療健康系 | 医療・健康関連の情報を提供している Web ページ |
| ニュース系 | 新聞社が提供しているニュースの Web ページ |
| 京都観光系 | 京都の観光情報を提供している Web ページ |

ゴリに属する Web ページをそれぞれ 25 ページずつ、合計 125 ページを収集した。ここから、3 人の被験者にそれぞれ各ページごとに 2 つの注目語を選択してもらい、合計 750 個の評価用の注目語を収集した。

以下、5.1 節では予備実験として、提案方式で良い検索結果が得られるための最適な検索語数を検証し、さらに、提案方式を適用することで検索結果の満足度を改善可能な場合がどの程度存在するかを検証する。次に、5.2 節では、5.1 節 (2) で得られた注目語のみでは十分に満足する検索結果が得られない場合において、提案方式を適用することでどの程度の満足度の改善が見込めるかを検証する。ここでは、検索エンジンとして Yahoo!API¹³⁾ を利用した。

5.1 予備実験

(1) 提案方式に最適な検索語数の検証

● 実験方法

検索語数を 1~5 個に調整して検索した結果の第 1 位の Web ページの満足度を評価した。検索語の選択方法は、提案方式のパラメータの影響を避けるため、最も直観的な指標である注目語からの距離 ($F_d(w, w_a)$) に基づき選択した。検索語数が 1 個の場合は、注目語のみでの検索を意味し、2 個目以降の検索語は注目語からの距離に近い順に選択した。評価値は 3 段階 (最低 0 点, 最高 2 点) とし、それぞれ、“注目語とまったく見当違いの内容、または検索結果が 0 件 (0 点)”, “注目語の意味と内容がやや関連する (1 点)”, “注目語の意味と内容がぴったり合っている (2 点)” を評価基準とした。

● 結果と考察

カテゴリごとに 3 人の評価値を平均した結果、および全カテゴリの結果の平均値を図 3 に示す。平均値は 3 語以上でほぼ安定していることから、本論文における提案方式の検索語数は、注目語を含めて 3 個とする。

31 周辺語を活用したクリック型 Web 検索システムの提案と評価

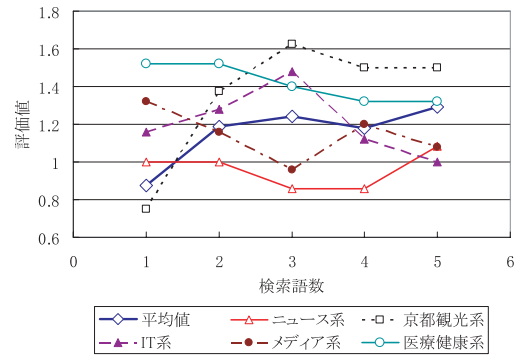


図 3 検索語数と評価値
Fig. 3 Number of keywords and score.

(2) 提案方式を適用する範囲の検証

● 実験方法

注目語のみで検索した結果の第 1 位の Web ページの満足度を評価した。評価値は 5 段階（最低 0 点～最高 4 点）とし、それぞれ，“検索意図とまったく見当違いのページ内容、または検索結果が 0 件（0 点）”，“検索意図とページ内容が部分的に合っている（1 点）”，“検索意図とページ内容の主題が合っている（2 点）”，“検索意図に対してページ内容に満足できるレベルの記述がある（3 点）”，“検索意図とページ内容がぴったり合っている（4 点）”とし，“4 点”は目的の情報が確実に推薦されたと思った場合のみ付けることを条件とした。このようにして得られた評価結果を、注目語のみの検索結果で 3 点以上の評価となったページと、2 点以下の評価となったページに分類した。2 点以下のページは多くのユーザが欲するような標準的な情報では満足しなかった検索結果を表しており、提案方式による改善が見込まれる範囲を表している。

● 結果と考察

結果を表 3 に示す。表 3 より、カテゴリにより多少のばらつきがあるものの、平均すると 50% 以上の場合において、注目語のみでは検索結果に満足しない結果となった。また、IT 系と京都観光系を比較すると大きく結果が異なった。この要因として、IT 系の語は、「オンラインチャージ」「オープンシステム」「フレームワーク」などのように、使われる範囲が限定されないと意味を限定しにくい語が多いことや、「ファイアウォール」などのようにユーザの知りたい範囲により内容や記述の難易度が大きく異なる語が

表 3 注目語のみによる検索の評価結果（括弧内の数値はサンプル数 150 件に対する割合を表す）

Table 3 Evaluation of search by a core word only.

| | 3 点以上（注目語のみの検索で満足） | 2 点以下（提案方式で改善が見込まれる範囲） |
|-------|--------------------|------------------------|
| IT 系 | 51 (34.0%) | 99 (66.0%) |
| メディア系 | 64 (42.7%) | 86 (57.3%) |
| 医療健康系 | 70 (46.7%) | 80 (53.3%) |
| ニュース系 | 70 (46.7%) | 80 (53.3%) |
| 京都観光系 | 95 (63.3%) | 45 (36.7%) |
| 平均 | 70 (46.7%) | 80 (53.3%) |

多いこと、さらに、それらの語が多く出現する Web ページであってもその説明がなく使われている場合が多いことから、検索結果に対するユーザの検索意図との乖離が起りやすいと考えられる。一方、京都観光系の語は著名な寺社仏閣などの名所名や歴史上の人物名に関する語が多く、その語を説明する内容は史実に基づく不変の内容がほとんどであるため、検索結果に対するユーザの検索意図との乖離が起りにくいと考えられる。

次の 5.2 節では、従来方式で 2 点以下となる場合に対する提案方式の有効性について評価を行う。

5.2 提案方式の評価

● 実験方法

提案方式と、注目語のみで検索した場合（以降、従来方式と呼ぶ）の検索結果の満足度を比較するため、3 人の評価者に対して 5.1 節 (2) と同様に 5 段階評価による主観評価実験を行った。まず従来方式の結果の評価を行い、続いて提案方式の結果の評価を行った。

従来方式の結果に対して提案方式の満足度が高い場合（“評価が上がった”）、下がった場合（“評価が下がった”）、変化しなかった場合（“変化なし”）をそれぞれカウントした。また、従来方式で 3 点以上の評価となった場合（“注目語のみで満足できた”）に対しても同様に、提案方式の満足度の変化をカウントした。3 人の被験者の評価を総計した結果をそれぞれ表 4 および表 5 に示す。 α の値は、学習データとして任意の 10 サイトをサンプルとして抽出し、式 (1) の第 1 項の平均値と第 2 項の平均値が同じになるように算出した ($\alpha = 0.06081$)。また、式 (7) の k の値は、周辺語が 3 語～5 語抽出できる程度になるように実験的に決定し、 $k = 20$ とした。ここで、式 (5) の $df(w)$ の

32 周辺語を活用したクリック型 Web 検索システムの提案と評価

表 4 注目語のみで検索する方式に対する提案方式の満足度の差（カッコ内の数値は「評価が上がった」「変化なし」「評価が下がった」の合計値に対する割合を示す）

Table 4 Evaluation of gain on proposed method.

| | 評価が上がった | 変化なし | 評価が下がった |
|-------|-------------|-------------|------------|
| IT 系 | 29 (29.3%) | 55 (55.6%) | 15 (15.1%) |
| メディア系 | 18 (21.0%) | 53 (61.6%) | 15 (17.4%) |
| 医療健康系 | 29 (36.3%) | 49 (61.2%) | 2 (2.5%) |
| ニュース系 | 30 (37.5%) | 41 (51.3%) | 9 (11.2%) |
| 京都観光系 | 35 (63.6%) | 20 (36.4%) | 0 (0%) |
| 合計 | 141 (35.3%) | 218 (54.5%) | 41 (10.2%) |

表 5 注目語のみの検索結果で満足できた場合に対する提案方式の満足度の差（カッコ内の数値は「評価が上がった」「変化なし」「評価が下がった」の合計値に対する割合を表す）

Table 5 Evaluation of gain on proposed method in the case of over 3 points by using conventional method.

| | 評価が上がった | 変化なし | 評価が下がった |
|-------|----------|------------|-------------|
| IT 系 | 0 (0%) | 13 (25.5%) | 38 (74.5%) |
| メディア系 | 0 (0%) | 7 (10.9%) | 57 (89.1%) |
| 医療健康系 | 0 (0%) | 21 (30.0%) | 49 (70.0%) |
| ニュース系 | 1 (1.4%) | 12 (17.1%) | 57 (81.5%) |
| 京都観光系 | 4 (4.2%) | 42 (44.2%) | 49 (51.6%) |
| 合計 | 5 (1.4%) | 95 (27.1%) | 250 (71.5%) |

算出に必要なドキュメント集合として、1994 年から 2006 年までの毎日新聞オンラインニュースの 135 万件の記事を利用した。

● 結果

表 4 より、従来方式で 2 点以下の場合においては、医療健康系および京都観光系のカテゴリにおいて提案方式が従来方式を大幅に上回る評価結果が得られた。特に、京都観光系のカテゴリにおいて、63.6%のケースで有効性が確認できた。その他の 3 つのカテゴリについても提案方式が上回っており、提案方式の有効性を確認できた。また、表 5 より、従来方式で 3 点以上の場合においては、京都観光系のカテゴリにおいて評価結果が均衡しているものの、すべてのカテゴリを総合的に見ると、71.5%のケースにおいて提案方式が注目語のみの検索結果を下回った。よって、従来方式の検索結果にユーザが満足できるケースにおいては、提案方式が上回る可能性が低いといえる。

● 考察

以上の結果より、提案方式と従来方式は単独では十分な満足度を実現できていないとい

える。ここで、表 4 で示したように従来方式で満足できない場合は提案方式で改善できること、および表 5 で示したように従来方式で満足しているケースにおいては、提案方式を適用しても効果が薄いことが観測できる。このことから、まず従来方式の結果をユーザに提示し、その結果満足しないりアクションが得られた場合には続いて提案方式の結果を提示するなど、提案方式と従来方式を組み合わせた実装方法が有効であると考えられる。これについては 5.4 節のシステム実装例で検証する。

また、個々の評価結果を観測すると、提案方式が有効に作用するケースを次の 3 つに分類できた。

ケース 1：注目語だけでは文章量が少なく、十分な情報が得られないケース

ケース 2：注目語だけでは法律の原文などの資料そのものが検索されてしまうケース

ケース 3：注目語の部分一致で検索され、注目語そのものを検索できないケース

それぞれのケースについて、提案方式が有効に作用した具体例を表 6 に示す。一方、提案方式が有効に作用しない事例として、関連する文書の種類が少ない場合において、周辺語を含む Web ページとして実験対象の Web ページが推薦されてしまうケースがあった。また、4 章で述べた提案方式の (Step2) で生成した複合語が、検索サイトの検索語として登録されていないものであったため、検索結果が 0 件となってしまったケースが見られた。前者については、URL の一致を見ることで解決可能であり、後者については、検索結果が 0 件となる時、複合語を生成前の状態に戻して再検索をすることで解決可能であると考えられる。

メディア系とニュース系のカテゴリで提案方式の有効性が顕著とならなかった原因として、ニュース系サイトに出現する語の多くは、厳密な定義が与えられていることがあげられる。たとえばニュース系のカテゴリにおいては、書籍のタイトルにも利用されるような語が多いことや、同じ語を含むような同様の内容のニュース記事が別のサイトにも掲載されている場合が多い。そのため、たとえば注目語だけでは書籍のオンライン販売のページが提示されることがある。このとき周辺語を組み合わせることで正しい情報を提示するページを提示できる場合が多くあり、その結果、提案方式の評価は上がる。また逆に、周辺語を使うと注目語を選択したページと同様のニュース記事が提示されてしまう場合が多くあり、このときの提案方式の評価は下がってしまう。これにより、特にニュース系のカテゴリにおいて提案方式の評価結果が上がった場合と下がった場合に大きく分かれてしまったと考えられる。同様のニュース記事が提案されてしまう場合に

33 周辺語を活用したクリック型 Web 検索システムの提案と評価

表 6 検索結果の具体例 (太字は抽出した注目語を表す)

Table 6 Examples of results.

| | | 従来方式 | 提案方式 |
|-------|---------|--|-------------------------------------|
| ケース 1 | 原文 | 「...今回は明治から昭和初期に活躍した「神坂雪佳」を紹介する。当時多くの芸術家が西欧へ...」 | |
| | 検索語 | 神坂雪佳 | 神坂雪佳, 明治, 昭和初期 |
| | 結果ページ概要 | 文章量が比較的少ないページ | 文章量が比較的多いページ |
| | ユーザ評価値 | 2 | 4 |
| | 原文 | 「... 質調査所によると、インドネシア・ジャワ島沖で九日午前零時(日本時間同二時)ごろ、...」 | |
| | 検索語 | ジャワ島沖 | ジャワ島沖, 九日午前零時, 日本時間 |
| | 結果ページ概要 | 要点だけを述べた比較的短いニュース記事 | 文章量が比較的多い詳しいニュース記事 |
| | ユーザ評価値 | 1 | 2 |
| | 原文 | 「... 尾, 京の吉野と並び称された大阪の名妓「夕霧」の墓がある清涼寺で追善法要が行われま...」 | |
| | 検索語 | 夕霧 | 夕霧, 京, 吉野 |
| | 結果ページ概要 | 文章量が比較的少ないページ | 文章量が比較的多いページ |
| | ユーザ評価値 | 1 | 2 |
| | 原文 | 「... 1855年) に造営されたものです。紫宸殿を始めとし、清涼殿、小御所、御学問所及び...」 | |
| | 検索語 | 紫宸殿 | 紫宸殿 学問所 清涼殿 |
| | 結果ページ概要 | 文章量と写真の枚数が比較的少ないページ | 文章量が多く、写真を多用したページ |
| | ユーザ評価値 | 3 | 4 |
| ケース 2 | 原文 | 「... 財政改革や格差問題などが争点。「従軍慰安婦」などをめぐる上田氏の歴史認識も議論と...」 | |
| | 検索語 | 従軍慰安婦 | 従軍慰安婦, 上田氏, 格差問題 |
| | 結果ページ概要 | このタイトルの本の目次と感想が書かれたページ | この問題に関連するニュース記事 |
| | ユーザ評価値 | 1 | 2 |
| | 原文 | 「... が米艦船などに給油活動するためのテロ対策特別措置法が11月1日で期限切れとなるこ...」 | |
| | 検索語 | テロ対策特別措置法 | テロ対策特別措置法, 米艦船, 給油活動 |
| | 結果ページ概要 | 法律の原文がそのまま記載されたページ | 法律に詳しくない人でも分かるように平易に法律の内容を説明しているページ |
| | ユーザ評価値 | 3 | 3(評価値は変わらず) |
| | 原文 | 「... が揺らいでいないか、と突き付ける内容の平和宣言を読み上げた。銃撃され死亡した伊藤...」 | |
| | 検索語 | 平和宣言 | 平和宣言, 内容, 銃撃 |
| | 結果ページ概要 | 宣言の原文がそのまま記載されたページ | 原文を引用しながら事件にも触れて著者の意見も含むページ |
| | ユーザ評価値 | 2 | 3 |
| ケース 3 | 原文 | 「... 爆発し、乗組員7人全員が死亡した。「宇宙の先生」計画は中断され、モーガンさんは学...」 | |
| | 検索語 | 宇宙の先生 | 宇宙の先生, 乗組員7人全員, モーガンさん |
| | 結果ページ概要 | 「宇宙」を含むページ | 事故に関連するニュースのページ |
| | ユーザ評価値 | 1 | 2 |

については、注目語前後の文字列に対する一致を検査しておくことで、解決可能と考えられる。

次の 5.3 節では、提案方式の有効性に寄与した個々の要素について検証する。

5.3 各要素の影響度評価

提案した周辺語重要度モデルの個々の要素について評価実験を行い、提案方式の有効性に寄与した要因を検証する。

(ア) $F_m(w)$ および $F_v(w)$ の評価

● 実験方法

$F_m(w)$ 要素と $F_v(w)$ 要素の有効性の評価を行った。 $F_m(w)$ の評価のための検索語として、注目語に加えて、形態素解析で「名詞-人名」「名詞-地名」と判定された周辺語を 2 語選択した。不足する場合は、注目語からの距離が近い語から順に選択した。注目語の近隣に人名および地名を含むという条件に該当する評価結果について、検索結果第 1 位の Web ページに対する評価値を集計した。一方、 $F_v(w)$ の評価のための検索語として、出現数の多い「」, 『』, および $\langle b \rangle \langle /b \rangle$ タグで囲まれた語を 2 語選択した。不足する場合は同様に、注目語からの距離が近い語から順に選択した。注目語の近隣に強調表現を含むという条件に該当する評価結果についても同様の評価値を集計した。なお、これらの条件に該当する評価結果の数が少ないため、ここでは 5 つのカテゴリの結果を総計し、地名優先の評価総数は 53, 人名優先の評価総数は 29, 強調表現優先の評価総数は 154 となっている。5.2 節と同様の手法で従来方式に対する評価結果の差を集計した。

● 結果と考察

結果を表 7 に示す。表 7 より、周辺語に地名を優先的に利用することで、評価が上がっており、地名を優先的に利用する効果は高いといえる。一方、人名を優先的に利用した場合は、ほとんど評価が変わっていないことから評価への影響が小さい。よって、地名を利用する場合と比較して人名を利用する場合の効果は薄いといえる。また、周辺語に強調表現の語を利用することで評価が上がっており、強調表現の語を優先的に利用する効果は高いといえる。

(イ) $F_c(w, w_a)/F_d(w, w_a)$ の評価

● 実験方法

$F_c(w, w_a)$ 要素と $F_d(w, w_a)$ 要素の有効性の評価を行った。検索語は注目語に加えて、 $F_r(w, w_a)$ の値が大きい順に周辺語を 2 語選択した。従来方式との評価結果との差につ

34 周辺語を活用したクリック型 Web 検索システムの提案と評価

表 7 人名優先と地名優先および強調表現優先の評価結果

Table 7 Priority evaluation of person's name, place name and emphasized word.

| | 評価が上がった | 変化なし | 評価が下がった |
|--------|---------|------|---------|
| 地名優先 | 12 | 36 | 5 |
| 人名優先 | 4 | 19 | 6 |
| 強調表現優先 | 55 | 66 | 33 |

表 8 共起語優先の評価結果

Table 8 Priority evaluation of cooccurrence words.

| | 評価が上がった | 変化なし | 評価が下がった | 注目語のみで満足できた |
|-------|---------|------|---------|-------------|
| IT 系 | 38 | 41 | 20 | 51 |
| メディア系 | 33 | 42 | 10 | 65 |
| 医療健康系 | 42 | 34 | 4 | 70 |
| ニュース系 | 23 | 35 | 11 | 81 |
| 京都観光系 | 34 | 37 | 0 | 79 |
| 合計 | 170 | 189 | 45 | 346 |

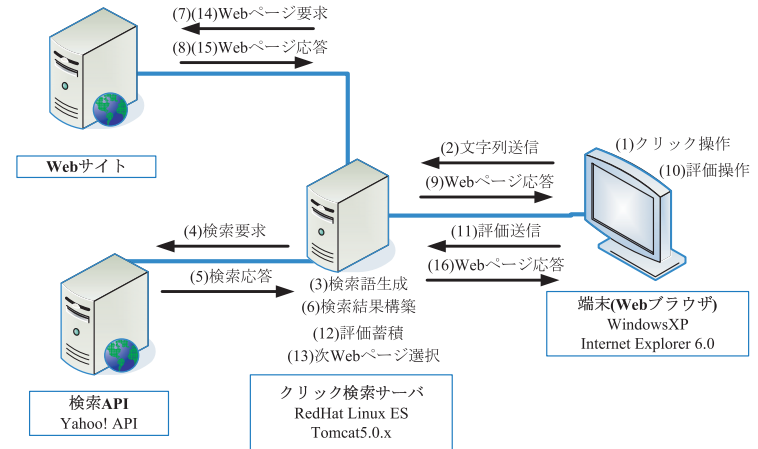


図 4 クリック検索システムの実装例

Fig. 4 Implementation example of click-based web search system.

いて、5.2 節と同様の評価手順により集計した。

● 結果と考察

結果を表 8 に示す。表 8 より、医療健康系および京都観光系サイトでは、従来方式を大幅に上回る結果となっており、従来方式で満足できなかった場合は、提案方式を利用することでほとんどの場合は満足度が上がっている。またその他 3 カテゴリについても、多くの場合、満足度が向上している。以上により、共起語を検索語に加える優先度を上げることは、検索結果の満足度向上に有効であるといえる。

5.4 システム実装例

提案方式を実装した検索システムを構築した。機能構成と動作シーケンスを図 4 に示す。本システムは、クリック検索サーバを中心として、Web ブラウザをユーザインタフェースとした端末、一般の Web サイト、および一般の検索エンジンが提供している検索 API サーバの 4 つの機能から構成される。ここで、ユーザがクリックした位置の前後の文字列を取得する機能を Web ブラウザのプラグインとして実装した。また、5.2 節の考察に従い、注目語のみの検索で得られた第 1 位の Web ページをユーザに第 1 番目に提示する Web ページとし、第 2 番目以降の検索結果として提案方式の検索結果を続けるように実装した。この実装方法により、表 4 で得られた提案方式が有効に作用する範囲を有効に活用し、2 クリッ

ク以内で満足する結果が得られる可能性が高くなる。さらに、クリック検索サーバが応答した Web ページの良否を評価する機能をクリック検索サーバに実装し、その評価用 GUI を端末の Web ブラウザのプラグインとして実装した。本評価機能を利用して、注目語に対するユーザが良いと判断した Web ページをクリック検索サーバに蓄積することができるようになる。この評価情報を利用することで、ソーシャルフィルタとして検索結果に反映する機能が利用できる。この機能は、評価情報がサービス開始時などの初期状態では存在しないことや、時間の経過にともない Web ページ自体が情報更新することなどに対応した情報の信頼性を維持する方策が必要である。この点については、提案方式と連携することで解決できる。システムの処理は、図 4 に記載している処理 (1) ~ (16) の手順で行う。

以下に各処理の概要を述べる。

(1) クリック操作

ユーザが、Web ブラウザ上の任意の語をポインタでクリックする。

(2) 文字列送信

端末の Web ブラウザが、ポインタ位置の前後の文字列を切り出してクリック検索サーバに送信する。

(3) 検索語生成

クリック検索サーバが、端末から受信した文字列を利用して注目語と周辺語を抽出し、周辺語重要度に基づき検索語を生成する。このとき、注目語に対する十分に信頼できる評価情報が蓄積されている場合は、その Web サイトにページを要求し、(8)へ進む。

(4) 検索要求

クリック検索サーバが、検索 API に対して検索語を入力して検索結果を要求する。このとき、(a) 注目語のみ、(b) 注目語と周辺語の組合せの 2 通りで要求する。

(5) 検索応答

検索 API が、(a)、(b)2 つの検索結果を応答する。

(6) 検索結果構築

クリック検索サーバが、2 つの検索結果を合成した検索結果リストを構築する。具体的には、(a) の検索結果の第 1 位を 1 番目とし、2 番目以降に (b) の結果をつなぐ。

(7) Web ページ要求

クリック検索サーバが、検索結果リストの 1 番目の URL に基づきページを Web サイトに要求する。

(8) Web ページ応答

Web サーバが、要求された Web ページをクリック検索サーバに応答する。

(9) Web ページ応答

クリック検索サーバが、検索結果として 1 番目の Web ページを端末に応答する。

(10) 評価操作

ユーザが、検索結果の Web ページを閲覧し、検索結果として満足する情報を含む Web ページかどうかを評価する。満足した場合は、閲覧は終了する。満足しない情報であった場合は、低い評価を付与することで、次の Web ページを要求する。

(11) 評価送信

端末の Web ブラウザが、ユーザの評価結果をクリック検索サーバに送信する。

(12) 評価蓄積

クリック検索サーバが、ユーザの評価情報を、提示した Web ページと紐付けて蓄積する。

(13) 次 Web ページ選択

ユーザの評価が低い評価であった場合、クリック検索サーバが、検索結果リストの次の Web ページを選択する。

表 9 実装例における検索結果に対する満足度の推定 (括弧内の数値はサンプル数 150 件に対する割合を表す)

Table 9 Estimation of satisfaction for search results.

| 1 クリック目 | 3 点以上 | 2 点以下 | |
|---------|------------|-------------|--------|
| 2 クリック目 | — | 3 点以上 | 改善幅 |
| IT 系 | 51 (34.0%) | 68 (45.3%) | +11.3% |
| メディア系 | 64 (42.7%) | 71 (47.3%) | +4.7% |
| 医療健康系 | 70 (46.7%) | 86 (57.3%) | +10.7% |
| ニュース系 | 70 (46.7%) | 79 (54.0%) | +6.0% |
| 京都観光系 | 95 (63.3%) | 114 (76.0%) | +12.7% |

(14) Web ページ要求

クリック検索サーバが、次の Web ページの URL に基づき、ページを Web サイトに要求する。

(15) Web ページ応答

Web サーバが、要求された Web ページをクリック検索サーバに応答する。

(16) Web ページ応答

クリック検索サーバが、検索結果として次の Web ページを端末に応答する。以降、(10)~(16)を繰り返す。

以上に示した実装方法における、2 クリック以内で得られる検索結果の Web ページに対する評価結果を表 9 に示す。1 クリック目で満足した場合と 2 クリック目で満足した場合について比較した。

2 クリック以内で検索意図に合致した Web ページを推薦できる確率は 45.3%~76.0%であり、改善幅は最大で 12.7%となった。注目語のみの検索結果と比較して満足度を大幅に改善できたといえる。

6. おわりに

本論文では、多くの情報家電端末の入力インターフェースとして利用されているリモコン型 UI を用いた Web 検索操作の簡略化を目標とし、クリック型検索手法を提案した。本手法は、ユーザがポインタで任意に指定した注目語について、システムが自動的に周辺語を補完して検索を行い、検索結果をユーザに提示することができる。

周辺語の重要度を決定するための周辺語重要度モデルについて、語自身の重要度と注目語との関連性の 2 点からなるモデルを提案し、検索結果に対する満足度を測定するための主観評価実験を行った。その結果、注目語のみを検索語とした検索結果では 50%以上の場合

で満足しないことが分かり、そのうち、提案方式を適用することで最大で 63.6%の検索クエリで満足度が向上することを示した。また、注目語のみの検索結果と提案方式の検索結果を組み合わせることで、2 クリック以内での検索結果に対する満足度が最大で 76.0%となることを示し、提案方式の有効性を確認した。さらに、周辺語重要度を算出するための個々の要素について評価を行い、地名と強調文字、および注目語に対する共起語を優先的に検索語に利用することで、検索結果に対する満足度が向上することを示した。

今後の課題として、周辺語重要度のモデルの各要素の精細化や要素の追加、ならびに複数の注目語を入力するなどの機能拡張により、検索結果に対するユーザの満足度のさらなる向上を目指すことがあげられる。また、技術の実用化に向けて、実際のリモコン型 UI を本システムに適用し、操作数などの評価実験を行う必要がある。

参 考 文 献

- 1) Wu, J., Uemukai, T. and Sugaya, F.: A User-Customizable Xwindow-like GUI System for Mobile Phones, *ACM SIGGRAPH 2006* (Poster) (2006).
- 2) OneStat.com. <http://www.rankstat.com/html/en/seo-news1-most-people-use-2-word-phrases-in-search-engines.html> (2007/9/20 アクセス)
- 3) 松尾 豊, 石塚 満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, *人工知能学会論文誌*, 17-3D, pp.217-223 (2002).
- 4) 若木裕美, 正田備也, 高須淳宏, 安達 淳: 検索語の曖昧性を解消するキーワードの提示手法, *DBSJ Letters*, Vol.4, No.2 (2005).
- 5) 木本晴夫: 日本語新聞記事からのキーワード自動抽出と重要度評価, *電子情報通信学会誌*, Vol.74-D-I, No.8, pp.556-266 (1991).
- 6) Edmundson, H.: New Methods in Automatic Abstracting, *J. ACM*, Vol.16, No.2, pp.264-285 (1969).
- 7) Luhn, H.P.: A statistical approach to mechanized encoding and searching of literary information, *IBM Journal of Research and Development*, Vol.1, No.4, pp.390-317 (1957).
- 8) 石谷康人, 鈴木 優, 布目光生: 連鎖検索と近傍検索に基づく Web コンテンツへの効率的なアクセス方法, 第 6 回 Web インテリジェンスとインタラクション研究会 (SIG-WI2), pp.31-36 (2006).
- 9) Cutting, D.R., Karger, D.R., Pedersen, J.O. and Turkey, J.W.: Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections, *Proc. 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.318-329 (1992).

- 10) 小林拓海, 佐藤大介, 三末和男, 田中二郎: Web 検索結果の概観提示による情報収集支援インタフェース, 第 19 回人工知能学会全国大会, No.3C3-03 (2005).
- 11) 水口 充, 梅本あずさ, 柴尾忠秀, 浦野直樹: 提示型ユーザインタフェースの実装と評価, *コンピュータソフトウェア*, Vol.18, No.1 (2001).
- 12) 日本語形態素解析器 ChaSen. <http://chasen-legacy.sourceforge.jp/> (2007/9/20 アクセス)
- 13) Yahoo! Japan ウェブ検索 Web サービス. <http://developer.yahoo.co.jp/search/> (2007/9/20 アクセス)
- 14) Salton, G. and McGill, M.J.: *Introduction to Modern Information Retrieval*, McGraw-Hill (1983).

(平成 20 年 3 月 20 日受付)

(平成 20 年 7 月 7 日採録)

(担当編集委員 有次 正義, DBWeb 2007 推薦論文)



服部 元 (正会員)

平成 8 年神戸大学工学部電気電子工学科卒業。平成 10 年同大学大学院修士課程修了。同年国際電信電話(株)(現 KDDI(株))入社。現在(株) KDDI 研究所知能メディアグループ研究員。この間、ネットワーク管理、高度交通システム、ソフトウェアエージェント、Web アプリケーションの研究開発に従事。平成 15 年電子情報通信学会学術奨励賞受賞。電子情報通信学会、日本データベース学会各会員。



原 隆浩 (正会員)

平成 7 年大阪大学工学部情報システム工学科卒業。平成 9 年同大学大学院工学研究科博士前期課程修了。同年同大学院工学研究科博士後期課程中退後、同大学院工学研究科情報システム工学専攻助手、平成 14 年同大学院情報科学研究科マルチメディア工学専攻助手、平成 16 年より同大学院情報科学研究科マルチメディア工学専攻准教授となり、現在に至る。工学博士。平成 8 年本学会山下記念研究賞受賞。平成 12 年電気通信普及財団テレコムシステム技術賞受賞。平成 15 年本学会研究開発奨励賞受賞。データベースシステム、分散処理に興味を持つ。IEEE, ACM, 電子情報通信学会、日本データベース学会の各会員。



滝嶋 康弘

昭和 61 年東京大学工学部電子工学科卒業。昭和 63 年同大学大学院修士課程修了。国際電信電話(株)(現 KDDI(株))入社。現在(株)KDDI 研究所知能メディアグループリーダー。この間、動画像の符号化方式、動画通信システム、情報理論の研究・開発に従事。映像情報メディア学会、画像電子学会会員。工学博士。



菅谷 史昭(正会員)

昭和 57 年東北大学工学部通信工学科卒業。昭和 59 年同大学大学院修士課程修了。同年国際電信電話(株)(現 KDDI(株))入社。平成 9 年より平成 14 年まで ATR 音声翻訳通信研究所に出向。平成 14 年 KDDI(株)復帰。現在(株)KDDI 研究所執行役員。この間、情報検索、eLearning、音声翻訳評価の研究開発に従事。平成 3 年電子情報通信学会学術奨励賞受賞。電子情報通信学会、日本音響学会各会員。工学博士。



西尾章治郎(フェロー)

昭和 50 年京都大学工学部数理工学科卒業。昭和 55 年同大学大学院工学研究科博士後期課程修了。工学博士。京都大学工学部助手、大阪大学基礎工学部および情報処理教育センター助教授、大阪大学大学院工学研究科情報システム工学専攻教授を経て、平成 14 年大阪大学大学院情報科学研究科マルチメディア工学専攻教授となり、現在に至る。平成 12 年より大阪大学サイバーメディアセンター長、平成 15 年より大阪大学大学院情報科学研究科長、その後平成 19 年より大阪大学理事・副学長に就任。この間、カナダ・ウォータールー大学、ビクトリア大学客員。データベース、マルチメディアシステムの研究に従事。現在、Data & Knowledge Engineering 等の論文誌編集委員。本会理事を歴任。電子情報通信学会フェローを含め、ACM、IEEE 等 8 学会の各会員。