DBSJ Journal, Vol.10, No.1
June 2011

# リアクションサーチ ： リアクションをクエリとするウェブ情報検索

## Search-by-Reaction: The Web Search Using Reader's Reaction in Twitter

荘司 慶行♥　田中 克己♦

Yoshiyuki SHOJI　Katsumi TANAKA

**本研究では，“泣ける”や“行ってみたい”など，発見した文書を読んだ際に読み手がどのように感じるかをクエリとして入力可能なウェブ情報検索を実現する．現状で，ウェブ上で”C 言語に関する分かりやすい文書”など，読み手がどう感じるかに基づいた検索を行った場合，本当に望んでいる文書を発見することは難しい．既存の検索エンジンは文書中に出現する語に基づきインデックスを作成するため，書き手が本文中に記述した語に検索結果が大きく左右され，また読み手がどう感じるかは本文中に現れないため，望んだ文書が検索結果に現れない．一方，Twitter や SNS に代表されるウェブ 2.0 サービスにおいては，ユーザが紹介されたウェブページに対して“泣ける” や“分かりやすい” など，コミュニケーションの一環としてリアクション（反応）をとっている場合がある．そこで，本研究ではウェブコミュニケーションデータに含まれるリアクションをウェブページと対応付けし，クエリとして入力可能にすることで，読み手がどう感じるかに基づく検索が可能なウェブ検索システムについて提案する．本論文では，提案するシステムを実現する手法を提案し，実装する．**

This paper proposes a novel web search method named “Search by Reaction.” It enables searchers to search by impression terms as a query, such as “tear-jerker,” “want-to-go” and so on. Existing web search models only focus on terms appeared in main text. Terms in documents respond to only writers' intent. It is not always true that the interesting text includes the term “interesting.” Thus they cannot accept queries based on how the reader feels after read the retrieved document; e.g. “the document about diary, which makes me funny.”

However, it is common that users of web 2.0 services recommend their friends web sites, and then they take reaction for the page. The proposed method focuses on those reactions as the readers' impressions of web sites, and uses them for web search. We propose a basic algorithm that enables the search like this, and implement a prototype of the web search engine based on it.

## 1. はじめに

ウェブ2.0が叫ばれてから5年が経過したが，近年のウェブの有りようは，かつてのウェブに対して大きく変貌している．知識文書が主体だったウェブページは多様性を増し，商業サイトやウェブコミュニケーションサービス，個人の日記など，ウェブ上のリソースは今までになく多種多様になってきている．また，コンピュータやネットワークの知識に明るくない人たちも気軽にインターネットを利用するようになり，ウェブ利用者は利用者層，利用者数ともに増してきている．このように多様性を増している現代のウェブにおいて，従来の検索エンジンでは不十分な点が出てきている．

数も多様性も増した現状のウェブの中で，ユーザが望んだ文書を発見することは容易ではない．ことさら近年増加した利用者の中で，コンピュータに詳しくないユーザは，検索エンジンの仕組みを理解しておらず，自分の持っている検索要求を検索エンジンに満足に伝えることができない．またこういったユーザは先行知識に乏しく，既存の検索モデルの要求するような文中語形式のクエリがうまく作成できない．こういったユーザでも，望んだ文書が発見可能な検索システムの必要性が高まっている．

こういったユーザが普段しているであろう検索モデルを考える上で，インターネット上から離れて，現実世界における検索行為について目を向ける．現実世界における文書検索の具体的な例として，詳しくない分野についての書籍を探す場合を挙げる．ある分野に詳しくないユーザが，書籍を紹介してもらう際，例として“マヌエルとマイクロフトという登場人物が登場するSF小説を紹介してほしい”や，“C言語に関して，ポインタという語を含む本を教えてほしい”というような聞き方をすることは稀である．登場人物名や具体的な内容など，本文中に含まれる要素に基づいた検索は，対象の文書の内容をある程度理解していなければならないため，現実世界ではあまり行われない．通常，このような検索を行う場合には，“読んだら面白いと感じる，SFの本を教えてほしい”や，“C言語に関する本で，わかりやすい本を探してほしい”というような聞き方をする．現実世界でこういった検索をする際には，本文の内容や含まれる語に基づく検索でなく，“わかりやすい”，“スカっとする”，“泣ける”など，文書を実際に読んだ際に読み手がどのように感じるかに基づいた検索をしている．

このような“読み手がどう感じるか”に基づくような検索は，現状のインターネット検索においては，満足に行えない．先の例のような検索をしようとして，従来の検索エンジンにクエリとして“C言語わかりやすい”と入力した場合，望むような検索結果は得られない．“わかりやすいC言語の本を探しているが，見つからない”という個人の日記や，“我が社のわかりやすいC言語の本を買ってください”という通信販売のページなどが検索上位に並ぶ一方，本当に必要なページは検索結果に現れない．このようなことが起こる原因は2点ある．一点目は，従来型の検索モデルの問題点に起因する．現状の検索エンジンでは，検索対象とする文書中に含まれた単語に着目した検索を行う．そのため，“わかりやすい”という語が一度も登場しないがわかりやすい文書があった際など，その文書を発見できず，検索結果の再現率が低下する．もうひとつの原因は，文書に現れている語は，あくまで書き手の意図によるもので，読み手がどう感じるかとの間に相違がある点である．書き手が面白いと思ったものが読み手にとって面白いとは限らないほか，商業的なサイトにおいては，ポジティブなイメージをもつ語を売り文句として記載している場合がある．そのため，読み手が面白いと感じない文書が検索結果に現れ，適合率が低下する．従来の検索エンジンの，文中語をクエリとする情報検索は，すでにある程度知っ

♥ 学生会員　京都大学社会情報学研究科 博士後期課程
shoji@dl.kuis.kyoto-u.ac.jp
♦ 正会員　京都大学社会情報学研究科 教授
tanaka@dl.kuis.kyoto-u.ac.jp

ている内容を掘り下げたい場合や，以前に一度読んだ文書を探している場合には有用であるが，新規文書の検索に向かない．この問題を解決するため，“読み手が実際に文書を読んだ際にどのように感じるか”をクエリとして入力可能な検索システムを提案する．

提案システムでは，あるウェブページを読んだ際，読み手がどのように感じるかに基づく検索を可能にするために，ウェブ2.0コンテンツに含まれるユーザのリアクション（反応）に着目する．Twitter[1]やSNSといったウェブコミュニケーションを目的とするウェブ2.0サービスの中では，コミュニケーションの一環として，ウェブページを紹介したり，されたりしている場合がある．そして，紹介されたウェブページに対して，ユーザが“おもしろい”や“泣ける”などといった反応（リアクション）をとっている場合がある．このリアクションは，“読んだ時にこう感じる”というものが，具体性を伴って記述されたものである．具体例として，猫の画像の掲載された“かわいい”ウェブサイトを紹介されたユーザたちは，“キュートだ”，“プリティ”，“抱きしめたい”といった文をリアクションとしてコミュニケーションサイト投稿している．これらウェブコミュニケーション中のリアクションを収集し，ウェブページと対応付けることで、

- （“かわいい”，“猫”）
- （“行ってみたい”，“イベント”）
- （“よくある”，“大学生活”）
- （“賛成”，“法案”）

という形式で，通常の検索エンジンにおけるクエリに加え，文書を読んだ際にとられるリアクションをクエリとした検索を可能にする．これにより，上述したような，その文書を実際に読んだ際に読み手がどのように感じるかに基づいた検索が可能になると考えられる．

本稿では，このように，“リアクションに含まれる読み手がどう感じたかを表す語”をクエリとして入力可能なウェブ検索システムについて提案する．また，このようなシステムを実現する手法を提案し，プロトタイプを実装する．2節で本稿の提案するリアクションサーチと実現手法について述べる．また，3節で詳細な実装について，4節で本研究と関連する先行研究について述べ，5節でまとめと今後の課題について述べる．

## 2.リアクションサーチ

通常のトピッククエリに加えて，リアクションをクエリとして入力することで，発見した文書を読んだ際にどのように感じるかに基づくウェブ検索ができる検索エンジンを実現するための手法について説明する．この検索システムの概要を図1に示す．システムへの入力は，読んだ際にどう感じるかを表すリアクションクエリと，何に関する文書かを表すトピッククエリである．これら二つのクエリは文字列として入力される．検索対象はすべてのウェブページであり，スコアリングにはウェブコミュニケーションのログおよびウェブページの内容を利用する．システムの出力は，クエリにマッチするウェブページの順位付きのリストである．ページの順位は，ページとリアクションクエリ，トピッククエリとの適応度によって決定される．

システムの受け付けるリアクションクエリは，読み手が文

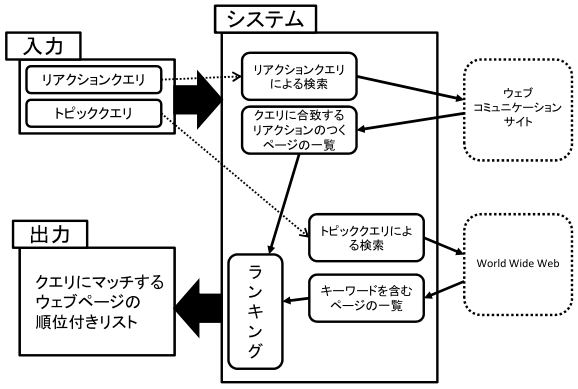


**図 1 ウェブ検索システムの概要**

Fig.1System outline

書を読んだ際にどう感じるかを表す語である．ここで言うリアクションとは，文書を読んだ際にどう感じたかがコミュニケーションを通じて記述されたものである．この記述は具体性を伴って自然言語で行われるため，表現の形態に多様性がある．例として，読み手が¥分かりやすい”と感じるような文書を読んだ際に実際にとられるリアクションは，

(1) “簡単だ”，“難しくない”などの換言表現
(2) “おかげで理解した”などの因果に基づく表現
(3) “へぇ！”，“なるほど”などの感嘆表現

など，さまざまである．提案するシステムでは，これらのうちどのような形式のリアクションでも，リアクションクエリとして入力可能である．また，“なるほど”というクエリ入力で“わかりやすい”と感じる文書を検索するなど，別の表現で記述されたリアクションに関しても検索可能にする．

### 2.1 検索モデル

すべての検索対象のウェブページ集合$P$, 全単語を$V$とした際，入力されるクエリ$q$は$q=(q_r \in q_t)$で表される．ただしこの際，$q_r \in V$, $q_t \in V$である．入力されたクエリ$q$を構成するリアクションクエリ$q_r$, トピッククエリ$q_t$について，リアクションサーチは$q_r$というリアクションのつく$q_t$に関するウェブページ$p$の集合をランクづけして返す．あるウェブページ$p$がリアクションクエリ$q_r$およびトピッククエリ$q_t$とどれだけ適合するかを表すランキング関数$rank(p, q_r, q_t)$を式(1)のように定義する．

$$rank(p, q_r, q_t) = scoteT(p, q_t) \cdot scoreR(p, q_t) \qquad (1)$$

ここで登場する関数$score_T(p, q_t)$はあるウェブページ$p$がトピッククエリ$q_t$に対してどれだけ適合するかの度合いであり，$score_R(p, q_r)$はウェブページ$p$がリアクションクエリ$q_r$に対してどれだけ適合するかの度合いである．これらの関数に関してはそれぞれ2.2項，2.3項で述べる．このようなモデルを実現するうえで，解決すべき技術的課題点を2点挙げる．

- リアクションクエリ$q_r$とウェブページ$p$との間の適合性計算を行う．この際，同じ意味を持つ別表現のリアクションに関しても検索対象とする．
- リアクションのひとつもついていないウェブページ$p0$に関してリアクションクエリ$q_r$との適合度を推定し検索対象化する．

本論文においては，特に前者に注力する．

[1] http://twitter.com/

### 2.2 トピッククエリに対するウェブページの適合性

あるページ$p$がトピッククエリ$q_t$に対してどれだけの適合性を持っているかの計算を，以下の式(2)で求める．本稿では，もっとも単純化したクエリ尤度モデル[1]を用いた．

$$score_T(p, q_t) = Pr(p \mid q_t) \propto Pr(p) \prod_{t \in q_t} Pr(t \mid M_p) \quad (2)$$

$Pr(q_t/M_p)$は，トピッククエリ$q_t$に含まれる語$t$が，ウェブページ$p$中に登場する回数を$tf_{t,p}$，ウェブページ$p$の長さを$L_p$としたとき，式(3)で推定される．

$$\Pr(q_t \mid M_p) = \prod_{t \in q_t} \Pr_{mle}(t \mid M_p) = \prod_{t \in q_t} \frac{tf_{t,p}}{L_p} \quad (3)$$

### 2.3 リアクションクエリに対するウェブページの適合性

あるウェブページ$p$についているリアクションが，どれだけリアクションクエリ$q_r$に対して適合性が高いかをページごとに取りまとめたスコア$score_R(p, q_r)$について説明する．リアクションとは，“ある文書を読んだ際にどう感じたか”が具体性を伴って自然言語で記述されたものであり，文書を読んだ際に感じたことそのものではない．そのため，わかりやすい文書に実際につくリアクションは，“わかりやすい”のほかに，“なるほど”，“へえ！”など，多彩な表現を伴う．そこで，リアクションクエリ$q_r$に対するウェブページ$p$の適合性を計算するためには，同一の感動を表すが別の表現で記述されたリアクションについて考慮する必要がある．ウェブページ$p$につくリアクション集合を$R(p)=\{r_1, \cdots, r_j\}$としたとき，それぞれ個別のリアクション$r \in R(p)$の$q_r$に対する適合度に関して，平均をとったものをページ$p$とリアクションクエリ$q_r$の適合度$scoreR(p, q_r)$として用いる．これを以下の式(4)で表す．

$$score_R(p, q_r) = \sum_{r \in R(p)} \frac{s\_r(r, q\_r)}{|R(p)|} \quad (4)$$

これら個別のリアクション$r$の入力されたリアクションクエリ$q_r$に対する適合度$s_r(r, q_r)$は，リアクション$r$を構成する複数の語$W(r)=\{w_1, w_2, \cdots w_j\}$についてそれぞれリアクションクエリ$qr$との適応度をとり，少なくとも一語が$q_r$と同義であるかを表す式(5)により合算することで決定する．

$$s_r(r, q_r) = 1 - \prod_{w_i \in W(r)} (1 - s_w(w_i)) \quad (5)$$

ここで，リアクション$r$を構成するそれぞれの語$W(r)=\{w_1, w_2, \cdots, w_i\}$について，その語が入力されたリアクションクエリqrとどれだけ近い意味を持つかについて考える．ひとつのウェブページに対して，複数のユーザから，複数のリアクションが付けられる場合がある．通常，”かわいい”というリアクションのつくページには，”かわいい”と同義のリアクションが付きやすいと考えられる．その一方で，同じ文書を読んだ場合でも，人によって，どのように感じるかは異なるため，同じページに対する複数のリアクションが，常に同種の感動を表すものだとは限らない．そこで，リアクションに含まれるある語$w_i$がリアクションクエリ$q_r$と同じ感動を表す語である度合い$s_w(w_i, q_r)$を，式(6)によって定義する．

$$s_w(w_i, q_r) = s_a(q_r, w_i) \cdot s_c(q_r, w_i) \quad (6)$$

式中の$s_a(q_r, w_i)$は$q_r$と$w_i$の観点の同一性を，$s_c(q_r, w_i)$は$q_r$が$w_i$とどれだけ特徴的に同じ対象に言及しているかをそれぞれ表す．ある語どうしが同一な観点に基づいており，なおかつ同じ対象に言及している場合，同じ感動に基づく別の表現のリアクションである可能性が高い．以下でこの2つのスコアに関して説明する．同一のウェブページに対して，複数の個人がそれぞれ別の観点に基づいたリアクションをとっている場合が考えられる．例として，動物の形をした家具を紹介するページを閲覧して，動物の形に注目して”かわいい”と感じる人もいれば，家具としての性質に注目して”便利そう”と感じる人もいる．そこで語wiがリアクションクエリ$q_r$と同一の観点である度合い$s_a(w_i, q_r)$を式(7)で定義する．

$$s_a(w_i, q_r) = \frac{P_{reaction}(q_r) \wedge P_{reaction}(w_i)}{|P_{reaction}(q_r)|} \quad (7)$$

$Preaction(w_i)$は語$w_i$を含んだリアクションが少なくともひとつ付いたウェブページの集合を返す関数である．ある語$w_i$がリアクションクエリ$q_r$と同一観点に基づく語であった場合，リアクションクエリ$q_r$を含むリアクションのついた他のウェブページに対するリアクションにもまんべんなく含まれていることが予想される．先の例において，“かわいい家具”へのリアクション，“かわいい菓子”へのリアクション両方に含まれている語は，同一観点に属すことが期待できる．“便利そう”という別の観点に基づくリアクションは前者の“かわいい家具”にしかつかない一方，“キュートだ”という同一観点に基づくリアクションは“かわいい家具”と“かわいい菓子”の両方につきうる．式$s_a(w_i, q_r)$はリアクションクエリ$q_r$を含むリアクションが付いたすべてのウェブページに，語$w_i$を含むリアクションがどれだけまんべんなく付いているかを表す．リアクションクエリ$q_r$がある語$w_i$とどれだけ特徴的に共起するかを表す尺度$s_c(w_i, q_r)$は，以下の式(8)で表される．

$$s_c(w_i, q_r) = \frac{|\{(r,p) \mid r \text{ includes } w_i, p \in P_{reaction}(q_r)\}|}{P_{reaction}(q_r)} \quad (8)$$

$Rreaction(w_i)$は語$w_i$を含むリアクションが少なくともひとつついたウェブページに対するリアクションすべての集合を返す関数である．ある語$w_i$がリアクションクエリ$q_r$とどの程度似た意味を持つかを判断するうえで，語$w_i$がリアクションクエリ$q_r$とどの程度特徴的に共起するかを利用する．先の例で，“かわいい”ウェブページに対するリアクションのみに含まれ，“かわいくない”ウェブページへのリアクションには含まれない語は，“かわいい”の同義語であることが期待できる．$s_c(w_i, q_r)$は，ある語$w_i$が，リアクションクエリqrを含むリアクションが付くようなウェブページにつくリアクションにどれだけ特徴的に登場するかを表す．以上の計算で，少なくともリアクションのひとつ以上ついたすべてのウェブページについて，付いたリアクションにリアクションクエリ$qr$が含まれていなくても，適合性$score_R(p, q_r)$を算出することができる．しかし，リアクションがひとつも付いていないウェブページ$p_0$があった場合，ページの内容にかかわらず，この適合性スコア$score_R(p_0, q_r)$は0となる．そこで，リアクションのついていないウェブページについて，$score_R$を推定する手法について，2.4項にて述べる．

### 2.4 リアクションのついていないウェブページのリアクションクエリに対する適合性の推定

インターネット上に公開されているすべてのウェブページに対して，ウェブコミュニケーションサイトにて紹介されリアクションがとられているウェブページは，ごく一部にすぎない．提案する検索システムでは，リアクションがひとつも付いていないウェブページについても検索可能にするため，リアクションクエリとリアクションのついていないページとの適合度についても定義する．内容の類似したウェブページは，読み手に対して同じ感動を与えると考えられる．そこで，リアクションのついているウェブページと付いていないウェブページの類似度をとることで，つくであろうリアクションを推定する．リアクションが付いており，なおかつトピッククエリ $q_t$にマッチするウェブページ集合 $P(q_t)$についてすべてのウェブページ $p \in P(q_t)$を連結した文書 $p_{qk}$について，各次元が語 $w_i$で，各次元の重みを語 $w_i$の $p_{qk}$における登場頻度 $tf(w_i, p_{qk})$とした特徴ベクトル $v_{pqk}$を作成する．そして，トピッククエリ $q_t$にマッチするウェブページのうち，リアクションのついていないウェブページ $p_0$に関しても，各次元を語 $w_i$，各次元の重みを語 $w_i$の $p_0$における登場頻度 $tf(w_i;p_0)$とする同様の特徴ベクトル $v_{p0}$を作成した．これら二つの特徴ベクトル $v_{pqk}$, $v_{p0}$について，式(9)の計算を行うことで，リアクションクエリ $q_r$とリアクションのついていないウェブページ $p_0$の適合度 $score_R(p_0, q_r)$を設定した．

$$score_R(p_0, q_r) = \cos(v_{P(q_t)}, v_{p_0}) \cdot min(score_R(p, q_r) / p \in p_{qk}) \quad (9)$$

ここで，リアクションクエリ $q_r$に適合するリアクションのついたすべての文書を結合してひとつの大きな文書にしたのは，リアクションのついた特定のページに極めて良く似た結果が検索上位に現れても情報が重複するだけで，検索結果の向上につながらないためである．また，リアクションクエリ $q_r$との適合度を計算する際に最小値をとっているのは，上位に似た結果が固まって表示されるのを避けるためである．

## 3.実装

実験提案するウェブ検索手法について実際に動作するウェブアプリケーションシステムを実装した．システムは2節で提案したアルゴリズムに則って動作する．ウェブブラウザからアクセス可能なウェブアプリケーションプログラムとして実装した．作成したウェブアプリケーションのスクリーンショットを図2に示す．二つ並んだテキストボックスの左側にリアクションクエリを，右側にトピッククエリをそれぞれ入力することで，上述のランキング関数にしたがって順序づけされた検索結果が得られる．この際，発見したウェブページのタイトル，URL，スニペットのほかに，実際にそのウェブページに対してなされたリアクションも表示される．

実装を行う上で，リアクションの収集はTwitterから行った．Twitter内の発言の中で，あるユーザが他人のURLを含む発言を引用しつつ，そのURLについてコメントしている形式のものを抽出し，コメント部分のみをリアクションとして利用した．収集したTwitter内の書き込みの具体例を図3に示す．この例において，検索の対象となるURLは“http://www.dl.kuis.kyoto-u.ac.jp”であり，このウェブページを読んだ際のリアクションが“こんなに面白いWebサイト見たことない www”である．また，リアクションのついていないWebページの検索にはYahooウェブ検索API[2]を利用した．これらのウェブページついてリアクションクエリに対する適合性の推定を行う際，特徴ベクトルの生成にスニペットを用いた．この際，リアクションのついていないページはYahooのランキングの上位100件について取得した．

図 2 実装されたシステムの結果画面

Fig.2 Screen shot of the implemented system


図 3 取得したリアクションの一例

Fig.3 An example of the scraped reaction

実際にアプリケーションとして実装する際，処理の流れは以下のようにした．

1. 前処理としてウェブコミュニケーションサイトからリアクションを収集し，URL とリアクションの対応付けを行い利用可能な状態にする
2. ユーザの入力したリアクションクエリとトピッククエリを受け取る
3. リアクションクエリを用いて，2.3 項の手法を用いて類似した意味の別表現を探し，単語のスコアを決定する
4. リアクションのついているウェブページの一覧に対して，トピッククエリで文中語による検索を行う
5. トピッククエリを用いて通常のウェブ検索を行う
6. 手順5で取得した結果について，リアクションのついていないものに関して，リアクションのついているものと類似度比較を行う
7. 手順3，6よりリアクションクエリとの適合度を，手順4，5よりトピッククエリとの適合度を算出し，ランキングを行い出力する

[2] http://developer.yahoo.co.jp/webapi/search/websearch/v2/websearch.html

実装上の都合として,検索結果の上限を3000件と定めた.この制限は手順 3 において課され，本文中に含まれるトピッククエリ語の登場回数による順位が3000位より下のウェブページに関しては，リアクションクエリとの適合度計算，検索結果への出力は行われない.

実装にはスクリプト言語である Ruby を使用し，作成したシステムは Debian Linux 上の apache2 により CGI プログラムとして動作する．リアクションおよびウェブページ本文の形態素解析には MeCab を,MeCab の辞書には IPA 辞書を用いた．この際，品詞情報などは考慮せず，いくつかのストップワードを除いたすべての語をスコア計算に利用した.

本稿における実装で用いたデータの規模について述べる.2010 年 11 月より 2010 年 12 月にかけて，Twitter ストリーミング API で無作為に取得した URL を含むツイートのうち，日本語で記述され，なおかつ先述した形式に則ったものをすべて利用した．収集した日本語ツイート総数は 30,134,604 であり，このうち 2,370,222 件がリアクションとして利用可能であった．リアクション中に含まれた URL の指すウェブページは，短縮 URL による重複を取り除いたうえで 1,286,752 件である．本実装において検索対象となるウェブページはこれらのウェブページと，外部の検索エンジンで取得したリアクションのついていないウェブページである．リアクションのついていないウェブページは検索時にトピッククエリにマッチする上位 100 件が取得される.

## 4.関連研究

本研究を行う上で，テーマ的，技術的な関連をもつ先行研究について説明する.

### 4.1 ソーシャルアノテーションの情報検索への活用

読み手がウェブページを閲覧した際にどう感じたかを付与する別の研究として，ソーシャルブックマークやタギングに代表されるソーシャルアノテーションが挙げられる．ソーシャルアノテーションの情報検索への利用は Heyman[2]，山家[3]らによって研究がなされている．本研究におけるリアクションも広義のソーシャルアノテーションの範疇に含まれるが，ソーシャルアノテーションサービスを通して付与されたタグとリアクションは性質が異なる．1 日当たりに付与されるアノテーションの量に関して，形式の固定されるこれらサービスのタグ付けに対して，Twitter をはじめとするウェブコミュニケーションから発見可能なリアクションは多い．また，アノテーションが行われるウェブページの質に関しても，わざわざタグ付けするまでもないような瑣末なページに対して，リアクションという形でならアノテーションが発見可能である．一方で，タグが単語からなり構造化が容易であるのに対し，リアクションは短文で表されることが多く，より多様な表現が許される.そのため，ユーザの生の声が反映されやすい反面，検索に利用する際には，2.3 で述べたような手法を用いて，別の表現を伴って印象が記述されている場合に表現の揺らぎを吸収する処理が必要となる.

### 4.2 評判情報検索

本研究においては，ウェブページを読んだ際にどう感じるかというユーザの記述をコミュニケーションサイトに含まれるリアクションからマイニングした．これと同様に，商品を購入，利用した際にどう感じるかというユーザの記述をブログや個人のウェブページなどから抽出する研究として，評判情報検索がある[4]．評判情報検索では，多くの場合，ある商品に関する評判がどのようなものであるかをウェブから収集することを目的とする場合が多く，また結果は観点と，ネガティブ/ポジティブによる極性からなる．評価表現をクエリとする商品検索の研究として杉木ら[5]の研究などがある.

### 4.2 センチメント分析と検索利用

文書の持つ感情情報を利用した検索手法として，センチメント検索[6]がある．ウェブページやニュース記事を内容から分析し，“明るい⇔暗い”，“承認⇔許否”などのいくつかの定められた軸ごとに感情の度合いを算出する[7]．ある文書の持つセンチメントの算出には文中の要素を利用しているため，現れるセンチメントは読み手ではなく書き手によって大きく左右される．この点が本稿における，読み手がどう感じるかに基づいたウェブ検索と大きく異なる.

## 5.今後の展望

本論文において提案した手法について，実装する過程で見出された課題と，それに対する今後の対応策について述べる.

今回はプロトタイプとして，もっともシンプルなランキングアルゴリズムを適用した．そのため，すべてのケースに適応できているわけではなく，一部で不適当なランキングが行われている．アルゴリズムの洗練化が必要である.

また,本稿においては,リアクションの収集もとを Twitter に限定し，特定の形式のもののみを用いた．これは本来ウェブ上に存在するリアクションのうちごく一部であるばかりか，不適当なものも含まれている．より精度の高いリアクションの収集が必要である．また，リアクションは Twitter に限らず，多くのウェブコミュニケーションから収集可能である．匿名掲示板 2ch[3]や Yahoo 掲示板[4]などへ収集対象を拡大することで，より多様な人の意見を反映する検索が可能になる.

類似したリアクションがトピックに対して非独立である場合，またページ間類似度がリアクションによって非独立である場合についても考慮する必要がある．今回の手法では，リアクションクエリと語の類似度は，トピッククエリとは独立に算出している．しかしながら，実際にはトピックによって同一の感動を表すリアクションは異なる．例として，カレーについて“おいしい”と“辛い”は類似したリアクションといえるが，トピックがケーキだった場合には“辛い”は“おいしい”と同義ではない．同様に，ページ間類似度についても，今回はリアクションの種類によらず，すべて単純なコサイン類似度に依っている．しかし検索に用いたリアクションの種類によっては，コサイン類似度では不適当な場合がある.例として，“簡潔だ”というリアクションに基づいた検索を行っている際，本文のコサイン類似度が高くても，長く冗長なページは結果としてふさわしくない．これら二つの点については今後対応してゆく必要性がある.

本稿で提案したシステムでは，クエリとして入力可能なリアクションは一種に限られ，検索の自由度が十分でない．複数のリアクションを入力可能であることが望まれる．この際，通常のトピック検索と異なり，単純なアンド検索，オア検索では不十分である.また，“賛否両論”や，“好みが分かれ

[3] http://www.2ch.net/
[4] http://messages.yahoo.co.jp/

る”など，リアクションのつき方の特徴による検索も考えられる．

先行研究として挙げたソーシャルアノテーションや，QAサイトに現れるユーザの入力したデータは，リアクションと似た性質を持っている．これらのデータを統合し利用可能にすることで，より多彩な意図を汲むことができ，より多くのページを網羅することが可能になる．

同じ文書を読んだ時にその文書から何を感じるかは，人によって大きく異なるため，検索のパーソナライゼーションが必要である．本稿においてはすべてのリアクションを同列に取り扱っているが，検索者と似た感性の持ち主のリアクションを重視するようなランキングにすることで，より読み手の立場に立った検索が可能になると考えられる．今回提案した手法では，リアクションのスコアリングをする際，使われている語や対象となっているウェブページのみを考慮し，発信者情報などは利用していない．ウェブページによっては一部の熱狂的なファンだけが好意的なリアクションをとっている場合など，バイアスがかかっていることが考えられる．ウェブコミュニケーションサイト上における発言者の関係性やユーザごとの嗜好などのソーシャルファクタを考慮することで，より万人受けするランキングにすることができると考えられる．

また，今回はリアクションのスコアリングに自然言語的な処理を施しておらず，否定語を含むリアクションが検索結果の精度を下げる要因となっている．リアクションのランキングに関しても，クエリとして入力されたリアクションとのマッチ度のみに基づいている．”とても”や，”少し”などの程度を表す表現があった際に，正しいランキングが行われない問題がある．スコアリング時に言語パターンを用いることで，否定語や程度などを考慮した検索が可能になる．

本稿においては，手法の提案を目的としているため，アルゴリズムはもっとも基礎的で単純な形式をとっている．実際にウェブ検索エンジンとして実用するためには，さらなる改良が必要である．

## 6.まとめ

本稿では，文書を読んだ際にどのように感じるかに基づいた読み手目線の検索を実現する，リアクション検索について提案を行った．提案手法ではウェブコミュニケーションデータを利用し，クエリとしてリアクションを入力可能にすることでこのような検索を可能にした．提案した手法について，実際にTwitterから取得したリアクションを用いることで作動するウェブアプリケーションを実際に作成した．

## [謝辞]

本研究の一部は，グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」，文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しいIT 基盤技術の研究」，計画研究「情報爆発に対応するコンテンツ融合と操作環境融合に関する研究」（研究代表者：田中克己，課題番号 1809041）によるものです．ここに記して感謝の意を表します．

## [文献]

[1] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schtze “Introduction to Information Retrieval,” Cambridge University Press (2008).

[2] P. Heymann, G. Koutrika and Hector Garcia{Molina: “Can social bookmarking improve web search?," Proceedings of the international conference on Web search and Web Data Mining, pp. 195 -- 206 (2008).

[3] Y. Yanbe, A. Jatowt, S. Nakamura and K. Tanaka: “Can Social Bookmarking Enhance Search in the Web?,” Pro-ceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2007), pp.107 -- 116 (2007).

[4] 乾孝司, 奥村学: “テキストを対象とした評価情報の分析に関する研究動向,” 自然言語処理, Vol.13, No.3, pp.201 -- 241 (2006).

[5] 杉木健二, 松原茂樹: “消費者の意見に基づく商品検索,” 情報処理学会論文誌,49, 7, pp.2598--2603 (2008).

[6] Na J.C. and Khoo C.S.G., Chan S. and Hamzah N.B.: “Sentiment-based search in digital libraries”, Digital Libraries, 2005. JCDL'05. Proceedings of the 5th ACM/IEEE-CS Joint Conference, pp.143--144 (2007).

[7] Tadahiko Kumamoto and Katsumi Tanaka: “Proposal of Impression Mining from News Articles,” Lecture Notes in Computer Science, Volume 3681/2005, pp. 901--910 (2005).

**荘司 慶行 Yoshiyuki SHOJI**

京都大学大学院 情報学研究科博士後期課程 在学中．2008年青山学院大学 理工学部情報テクノロジー学科卒業．2010年青山学院大学大学院 社会情報学研究科社会情報学専攻修士課程修了．ウェブ検索，ソーシャルサーチ，ウェブコミュニケーションからの知識抽出の研究に従事．日本データベース学会 学生会員．

**田中 克己 Katsumi TANAKA**

京都大学大学院情報学研究科社会情報学専攻教授．1976年京都大学大学院博士修士課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理，ウェブ検索の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，情報処理学会，日本データベース学会各会員．