

# オンラインニュースの有識性による評価

本多 一弥<sup>†</sup> 北須賀輝明<sup>†</sup> 有次 正義<sup>†</sup>

<sup>†</sup> 熊本大学自然科学研究科 〒 860-8555 熊本県熊本市黒髪 2-39-1

E-mail: †honda@dbms.cs.kumamoto-u.ac.jp, ††{kitasuka,aritsugi}@cs.kumamoto-u.ac.jp

**あらまし** Webによる情報検索，特にニュース検索によって我々が情報を得る機会が多い．本稿では，検索に不慣れたユーザの意図に沿い，且つ十分な内容をもつオンラインニュースを選択する指標として有識性を提案する．オンラインニュースを，クエリワード，記事の容量，他記事との類似度，更新時刻，元の検索結果における関連度による順位を数値化した有識性により評価する．より内容が多い記事ならば，トピックに関する情報が詳細で有識性も高く，また他記事と内容が類似する，または最近の記事であると，注目される話題を扱う記事と考えられる．ニュース検索結果を取得し，各要素を計算した値でリランキングし，実際に各要素を用いた有識性でよりよいニュースが取得できるか実験した．結果，有識性を用いてニュースを取得することは有効と考えられる．

**キーワード** ニュース検索，ランキング，Web，有識性

Kazuya HONDA<sup>†</sup>, Teruaki KITASUKA<sup>†</sup>, and Masayoshi ARITSUGI<sup>†</sup>

<sup>†</sup> Graduate School of Science and Technology, Kumamoto University

2-39-1 Kurokami, Kumamoto, Kumamoto 860-8555, Japan

E-mail: †honda@dbms.cs.kumamoto-u.ac.jp, ††{kitasuka,aritsugi}@cs.kumamoto-u.ac.jp

## 1. はじめに

現在インターネットを用いたオンライン上での情報検索は最も頻繁に行われるものの一つであり，オンラインニュースは様々な情報を手軽に取得する上で有用なツールである．オンラインニュースの利点として，リアルタイム性が高いことや，ユーザの興味による内容選択権があること，関連する内容をすぐに調べられることなどが挙げられる．また他の Web 情報に比べ，新聞社などの組織が情報提供するため一定の信頼性を持ち，同一トピックに関する記事が複数の組織によって提供されることから，様々な視点や考えを得ることができる．例えばスポーツであれば，相対する各チームの側にたった記事を読むことができる．

その一方で，ニュース検索サイトを用いることが不利益となる点もある．一度に多くの情報が提供されるため，ユーザが情報を取捨選択しなければならない．しかし，検索結果のページでは記事のタイトルとスニペットしか確認できず，より検索の意図に沿った記事かを判断することは困難だと考えられる．さらに検索時のクエリが曖昧な単語である場合，同音異義の単語に関する内容の記事も混在する場合があり，選択時に余計な手間がかかることや，誤った情報を得るという問題もある．例えば，「フィギュア」というクエリで検索した場合，ユーザがフィギュアスケートを意図して検索していても，人形のフィギュア

に関する記事が提供される可能性もある．

検索に不慣れたユーザが意図に沿った有益な情報を取得する上で，選択するための指標をたてることは重要である．ここで述べる有益な情報とは，ユーザの検索意図に沿っており，含まれる情報，具体的にはテキストが多く，さらに時間的にも適した情報とする．本稿では，この有益な情報を得る指標を有識性として定め，オンラインニュースにおける有識性を評価する．有識性は記事ページの単語量，類似度，更新時刻，クエリワード，元の検索順位に基づいて定量的に表す．ニュース検索結果を取得し，有識性の数値を基にリランキングすることで，有識性を用いたオンラインニュース評価がニュースの取得に有用であるかを考察する．

構成は以下の通りである．まず2章では，関連研究について述べる．3章では，本研究で提案する有識性に関して述べる．4章では，有識性を用いた実験を行い，考察する．最後にまとめと今後の課題に関して述べる．

## 2. 関連研究

### 2.1 情報信憑性 (Information Credibility)

Web の情報を用いるうえで，その情報が信頼できる情報かをまず確認することが重要である．情報の有識性を考える上で，その情報をどれほど信じていることができるかを示す情報信憑性に関する研究は重要な先行研究である．

情報信憑性の定義については既にいくつかの研究がなされており、例えば Tseng と Fogg [1] は信憑性を believability と定義し、信頼される人間は信用できる人間であるように、信頼される信用できる情報であるとした。また別の研究 [2] では、51 項目の Web に関するアンケートを 1400 人以上に対して行い、情報を提供する Web サイトの信憑性評価のポイントとして現実世界性 (Real-world feel)、使いやすさ (Ease of use)、専門性 (Expertise)、信頼性 (Trustworthiness)、オーダーメイド性 (Tailoring)、商業性 (Commercial implication)、そしてアマチュア性 (Amateurism) の 7 点をあげ、それらから人々の情報への認識の良悪を決定づけているという結論を導いた。

情報信憑性を高める方法として、加藤ら [3] は二つの方法を提案している。一つ目は情報を作成、発信する過程で品質の高い情報を発信できるプロセスを提案・構築する方法である。Web 以外の新聞や書籍等のメディアでは、情報を作成した後、幾度ものチェックを経てから発表となる。例えば新聞の場合<sup>(注1)</sup>、記者が情報収集後に記事を書き、支局長や支局デスクと呼ばれる人に記事として十分な内容であるかチェックをうけ、誤字訂正や内容補正、必要ならば再取材等を行い再びチェックをうける。次に本社のデスクに回され再度チェックをされ、その後記事のレイアウトへとまわされる。現在 Web 上で情報発信の際に、このようなチェック体制が徹底されているとは言い難く、特にブログや Twitter 等では自由に情報が発信できてしまう。Web 上でも新聞のようなチェック体制を提案・実施できれば、Web に存在し流通する情報全体の質の向上が可能となり、情報信憑性の向上にもつながると考えられる。

二つ目は、情報収集する際に情報の評価を行い、ユーザの必要性に応じて取捨選択のフィルタリングを行う方法である。この方法を実現することでユーザはより信憑性の高い情報、より価値のある情報を得ることが可能となり、取捨選択の基準となる指標を提供できればユーザの選択を補助・支援することが可能である。フィルタリングとしては、Web ページのデザインやレイアウト、誤字脱字等から評価する方法や、Fogg ら [2] が提案した要素を用いた評価システムを考案する等の手法が考えられる。

## 2.2 情報の評価

加藤ら [3] は、Web ページの情報の信頼性評価に関する要素として「皮相」、「意味」、「評判」、「発信者」を提案している。「皮相」は情報の見た目に関する要素であり、デザインや文体、単語に関する頻度分布等で表す。「意味」は情報の内容に関する要素であり、文構造や話題、記された情報が一般的であるか等を示す。次の「評判」は他ユーザからの評価に関する要素で、扱う情報、発信者、メディアの評判を考える。Google の PageRank も「評判」に含まれる。最後の「発信者」の要素は情報提供者の評価であり、信頼できる情報を発信しているかを発信者の履歴等から評価する。

Rieh ら [4] は、インターネットユーザの行動履歴を被験者か

ら提出してもらい、サイトや情報の種類別に、ユーザが何を基準に信憑性を判断しているのかを分析している。その研究の中ではオンラインニュースにおいては現在の風潮を捉えているかが重要であり、またニュースを含む多くの Web 情報において、ユーザは自分が良く知るサイトの情報を好んで閲覧する傾向が得られている。さらに、ユーザが情報の信頼性を確かめるには、情報の提供者や責任者を確認する機会が多いことも示されている。

分野やメディアを限定し評価する研究も行われている。Amin ら [5] は情報信憑性を評価する対象を、博物館のサイトや絵画に関するサイトやブログ等の、文化的内容の Web ページに限定して実験・評価を行った。12 の文化的サイトを参加者によって評価してもらい、評価結果を付加することで情報の選択時に有効かを実験している。サイトの信憑性は「信頼性 (中立的で有益であるか)」、「完全性 (不足なく情報が得られるか)」、「善意 (広告や他分野の情報がないか)」、「専門性 (分野で主流であるか)」の 4 項目を 5 段階で評価し、参加者自身の情報への既存知識も加味している。この研究では情報信憑性の評価結果を付加することで、ユーザがより自信をもって情報選択を行うことが可能となることを示した。

Nagura ら [6] は Web 上のニュース記事を対象とし、その信頼度を判定する方法を提案した。その指標としては「共通性」、「一貫性」、「信頼度」の 3 項目で評価している。「共通性」は複数の配信元間で、同イベントに関する記事が近い期間内に配信されるかで計算し、「一貫性」については“100 人の乗客”等の数量表現がある場合、他配信元の記事と比較し一致するかどうかで計算した。「信頼度」については、記事本文中で“〜かもしれない”といった推測や“〜という”等の伝聞表現を 4 段階でスコアリングし計算している。実験では人が評価した場合と提案手法を比較して 7 割ほどの一致が確認されている。

このように Web ページの情報評価に関して様々な研究がなされている。我々が提案する有識性は、この情報信憑性の一側面といえる。情報信憑性を構成する要素の一つである情報内容に注目し、Web ページから抽出した要素から計算された有識性は、ニュース検索においてユーザが必要とする情報かどうかを判断する指標の一つになり得ると考えている。

## 2.3 TextRank

Mihalcea と Tarau [7] は Google の PageRank を基に、自然言語テキストからキーワードや要約を抽出する手法である TextRank を提案している。TextRank はテキストから作成したグラフをベースとしたランキングアルゴリズムである。グラフを構成する頂点にはテキストに含まれる単語を用い、頂点をつなぐ辺はテキスト内での単語間の関係を用いる。単語間の関係とは、テキストを自然言語処理した際に、単語がテキストに現れた順番に並ぶが、この時単語同士が隣り合うことを示す。本研究では単語列において、前後に隣り合う単語同士を辺でつなぎ、グラフを作成する。図 1 にテキストをグラフ化した例を示す。

Mihalcea と Tarau はこのようなグラフをベースに、テキストを最適に表現するキーワードや文章の抽出が、TextRank に

(注1)：中日新聞「新聞ができるまで ~編集編~」,  
[http://www.chunichi.co.jp/nie/make/\(2012/3/13](http://www.chunichi.co.jp/nie/make/(2012/3/13) 現在)

『十和田市のホテルで16人食中毒（青森県）』  
 十和田市のホテルで食事をした2グループ16人が  
 ノロウイルスによる食中毒になった。東三番町の十和  
 田シティホテルで、7人からノロウイルスが検出され  
 た。県はホテルの食事が原因と断定し調理施設を  
 21日まで営業停止処分にした。快方に向かっている。

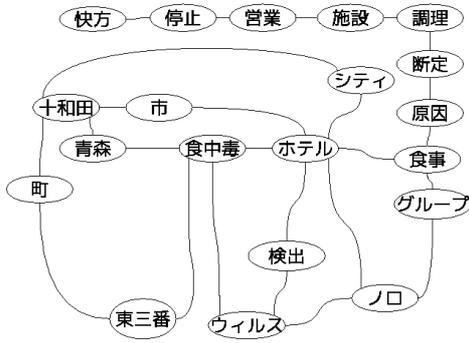


図1 テキストのグラフ化

よって正確に行われることを示し、またテキストからグラフを作成するため、深い言語知識や大規模なコーパスを必要とせず、さらに多言語、他分野にも応用可能である高い汎用性を示した。単語  $V_i$  の TextRank は、式 (1) のスコア  $S(V_i)$  で表される。

$$S(V_i) = (1 - d) + d \sum_{V_j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \quad (1)$$

$d$  は減衰係数であり、0.85 を用いている。  $In(V_i)$ 、  $Out(V_j)$  はそれぞれ、頂点  $V_i$ 、  $V_j$  が関係する単語群である。これは TextRank のモデルとなっている PageRank が有向グラフを対象として作成されているためである。今回 TextRank を用いる自然言語テキストから作成したグラフは、リンク関係のない無向グラフであるが、リンクが入る数と出る数が同等である有向グラフとも見ることができ、式 (1) は適用可能である。

PageRank では、他ページへ重複したリンクをもつことはほとんどないため、重みなしのグラフで定義されている。しかし TextRank は自然言語テキストから作成されたグラフであるため、特定の単語ペアが、他の単語ペアに比べて頻出する場合も考えられる。したがって二頂点、  $V_i$  と  $V_j$  をつなぐリンクに対して重み付けし重みを  $w_{ij}$  と表す。TextRank の式 (1) に重み付けを加味した式 (2) を次に示す。

$$WS(V_i) = (1 - d) + d \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \quad (2)$$

式 (2) を用いて、グラフの頂点である単語の初期値として 1 を与えてから、各頂点の最終スコアが定まるまで計算が行われる。最終スコアに収束するための条件としては、繰り返しの実行回数を制限するか、スコアの閾値を定める。最終スコアが定まった後は、スコアを基に逆順にソートされ、後の処理に利用するため保持される。論文 [7] ではこのスコアリングされた単語群からスコアが高い順に三分の一の単語を、キーワード群と

して定めている。

本稿では、オンラインニュースの有識性評価のため、TextRank を記事間の類似性計算に用いる値、また記事における検索クエリの値に用いる。またスコアリングされた単語群は全て評価に用いている。

### 3. 提案手法

本稿では、よりよいニュースを取得するための指標として有識性を定める。取得したいニュースとは、ユーザのニュース検索意図に沿った情報であり、且つその記事にユーザが知りたいことが十分に含まれているものと定義する。

#### 3.1 有識性

指標となる有識性がオンラインニュースから得られる要素を基に計算、ニュースをリランキングし、有効かどうかを検討する。本研究では次に示す 5 つの要素で有識性を構成する。

- (1) 単語量
- (2) 時間的側面
- (3) 類似度
- (4) 元検索順位
- (5) クエリ TextRank

各要素を用いることでよりよいニュースの複数の側面を定量化する。単語量は情報を伝えるテキストが十分であるか、時間的側面は検索クエリにおけるそのニュースがタイムリーであるか、類似度は注目されている内容であるか、元検索順位は一般に閲覧されるニュースかどうか、そしてクエリ TextRank はクエリワードがそのニュースの主題であることを示すと考えられる。

クエリに関する十分なテキストを含む記事であれば、ユーザが意図する情報を含むと思われ、またタイムリーで、他記事と類似する内容をもち、元検索順位が上位ならば、話題に合った内容と考えられる。

##### 3.1.1 単語量 ( $word_p$ )

単語量は記事テキストに含まれる非重複単語数と全単語数を考慮して求める。ここで述べる非重複単語とは、テキストを単語群として見た時、完全一致、または他の単語の一部として一致することがない単語のことをさす。例えば、「司会」と「司会者」という単語があった場合、「司会」は「司会者」の一部と一致しているため、「司会者」が非重複単語となる。

本文の長さ、すなわち文字数ではなく単語数を用いることで、単純に本文が冗長で同様な単語を繰り返しているニュースよりも、多種多様な内容を含むニュースが選択可能となると考えられる。また、全単語における非重複単語の割合は、非重複単語数が同じ値でも、単語の母数が少ない記事が、より重複しない内容を記事に含んでいると考えられるからである。

各ニュース記事において非重複単語を定めた後、式 (3) を用いて各記事の  $word_p$  を定める。

$$word_p = \frac{\text{非重複単語数}}{\text{最大非重複単語数}} + \frac{\text{非重複単語数}}{\text{全単語数}} \quad (3)$$

最大非重複単語数は、検索結果より取得した記事のうち、もっとも非重複単語が多い記事の数である。また非重複単語数と全

表 1  $time_p$  スコア時間対応表

時差	$time_p$ スコア
0 時間以上 1 時間未満	1. 0
1 時間以上 2 時間未満	0. 9
2 時間以上 4 時間未満	0. 8
4 時間以上 8 時間未満	0. 7
8 時間以上 16 時間未満	0. 6
16 時間以上 32 時間未満	0. 5
32 時間以上 64 時間未満	0. 4
64 時間以上 128 時間未満	0. 3
128 時間以上 256 時間未満	0. 2
256 時間以上 512 時間未満	0. 1
512 時間以上	0. 0

単語数は計算対象となっている記事のものである。

### 3.1.2 時間的側面 ( $time_p$ )

取得した最終更新時刻 (アップ時刻) が現時刻からどのくらい前のもので計算する。  $time_p$  は現時刻と最終更新時刻の時差から表 1 に則って定める。

### 3.1.3 類似度 ( $sim_p$ )

記事の形容詞と名詞のみを残し、残した単語  $V_i$  の TextRank  $TR(V_i)$  を、式 (4) で求める。

$$TR(V_i) = (1 - d) + d \sum_{V_j \in Rel(V_i)} \frac{w_{ji}}{\sum_{V_k \in Rel(V_j)} w_{jk}} TR(V_j) \quad (4)$$

$Rel(V_i)$  は単語  $V_i$  と関連している単語群である。  $Rel(V_i)$  は、記事テキストの形容詞と名詞のみを残した単語列において、  $V_i$  の直前か直後にある単語からなる集合と定義する。  $w_{ji}$  は、単語  $V_j$  と  $V_i$  の関係の重みであり、  $V_j$  と  $V_i$  が単語列で隣り合った回数が値となる。

計算は値が閾値を下回るか、試行回数が制限回数に達するまで繰り返される。各単語には初期値として 1 が与えられている。本研究では Mihalcea と Tarau [7] が用いた値と同様に閾値を 0.0001、試行の制限回数は 30 回とする。

各記事における単語の TextRank 計算後、その値を基に、二つの記事間のコサイン類似度を式 (5) で求める。

$$Similarity = \frac{Vec1 \cdot Vec2}{|Vec1| \times |Vec2|} \quad (5)$$

$Vec1$ ,  $Vec2$  はそれぞれ計算対象となった記事の単語群の TextRank で構成されるベクトルである。各記事ごとに他の全ての記事とコサイン類似度を求め、その平均値を計算する。例えばニュース記事を 50 件取得した場合、各記事について他の 49 の記事とのコサイン類似度を求め、その平均値を求める。そして全記事で最高コサイン類似度平均値で各記事の平均類似度を割った値を、各記事の  $sim_p$  と定義する。式 (6) に定義式を示す。

$$sim_p = \frac{\sum_{i=1}^{N-1} Similarity_i}{(N-1) \cdot Similarity_{max}} \quad (6)$$

$N$  は取得されたニュース記事数、  $Similarity_i$  はニュース記事ページ  $i$  のコサイン類似度、  $Similarity_{max}$  は、全記事のコサイン類似度平均値のうち、最も大きい値である。

### 3.1.4 元検索順位 ( $rank_p$ )

元の検索順位から値を決定する。今回ニュース検索には Google ニュースを用いている。例えばニュース記事を 50 件取得した場合、元 1 位から順に 1, 0.98, 0.96, ..., 0.02 と値を割り振る。

### 3.1.5 クエリ TextRank( $que_p$ )

各記事で検索クエリワード  $V_{que}$  の TextRank を求める。全記事のクエリワードの TextRank 値のうち最も大きい値を探し、  $TR_{max}(V_{que})$  とする。次に各記事のクエリワード TextRank を最大値で割ったものを  $que_p$  とする。  $que_p$  は式 (7) で表される。

$$que_p = \frac{TR(V_{que})}{TR_{max}(V_{que})} \quad (7)$$

クエリワードが二つ以上の場合や、クエリワードによっては TextRank の値を取得できない問題がある。本実験では、取得成功した場合のみを扱っている。

### 3.1.6 有識性

以上列挙した 5 つの要素を用いて、各ニュース記事の有識性を計算する。計算は式 (8) で行う。

$$\text{有識性} = word_p + sim_p + time_p + rank_p + que_p \quad (8)$$

以上で得た値を基に検索結果をリランキングし、有識性による評価の有用性を考察する。

## 3.2 リランキング

ニュース検索で得られた各記事には、先に述べた各要素を用いた有識性の値が与えられている。本研究で提案するシステムでは、次に示す手順で取得したニュース検索結果をリランキングする。

- 1 システムをスタートし、検索クエリを入力。
- 2 ニュース検索結果ページを取得。
- 3 検索結果から、各ニュースの URL 取得 (検索順位も取得)。
- 4 各ニュースの URL からソースページを取得。
- 5 各ニュース更新時刻取得。失敗時はソースから時間表現抽出。
- 6 各ソースページからタイトルと本文を抽出。
- 7 本文とタイトルから非重複単語群作成、不適な単語除去。
- 8 本文とタイトルを形態素解析、単語群作成。不適な単語除去。
- 9 各ページごとに式 (3) を用いて単語量計算。
- 10 各ページ単語群の単語ごとに式 (4) で TextRank 計算。
- 11 TextRank を基に各ページの類似度を式 (5), (6) で計算。
- 12 検索時刻と更新時刻の時差から各ページの  $time_p$  を導出。
- 13 各ページの元検索順位に基づいて  $rank_p$  を導出。
- 14 各要素から、ページごとに式 (8) を用いて有識性を計算。
- 15 有識性に基づいてリランキング。
- 16 リランキング結果出力。

手順 8 で作成される各記事ごとの単語群は、名詞と形容詞から構成される。それ以外の品詞の単語はフィルタリングによって除去される。

## 4. 実験

### 4.1 実験準備

実験は、Google ニュース検索エンジンで各クエリを検索した場合に表示された関連度順におけるオンラインニュース記事上位最大 50 件を対象とする。本実験で用いたクエリは「センター」、「バルサ」、「箱根」である。検索に不慣れなユーザが、テレビやヘッドラインニュースで話題となっている事柄を調べるためにニュース検索することを想定し、クエリは話題に関して具体性に欠ける単語を選択した。

またニュース検索を行った日時はそれぞれ、「センター」が 2012 年 1 月 17 日でセンター試験の翌週、「バルサ」が 2011 年 12 月 19 日でクラブワールド杯決勝戦翌日、「箱根」が 2012 年 1 月 5 日で箱根駅伝閉幕の 2 日後と、それぞれのクエリがニュース検索で話題となっているであろう時期に検索を行った。

提案した有識性の目的は、オンラインニュース検索において、よりよい記事を取得することである。よりよい記事とは、ユーザの検索意図に合致し、十分なテキストを含んだ記事をいう。十分なテキストとは、その記事を読むことで検索の意図を満たすテキストであることをさす。また、取得するとは、標準的な検索エンジンの表示において、1 ページ目の表示結果に掲載される上位 10 件に含まれることとする。Google や Yahoo! といった一般的な検索エンジンであれば、検索結果は 10 件ずつ標準設定で表示されるので、1 ページ目に掲載されれば閲覧される可能性は高いと思われる。

### 4.2 実験結果

以下に実験結果の一部を示す。なお記述を簡略化するため各値は小数点以下 3 桁までで示し、以降は四捨五入する。また表内で現順位としている順位は有識性によるリランキング後の順位である。

表 2 「センター」における結果

現順位	元順位	有識性	$word_p$	$sim_p$	$time_p$	$rank_p$	$que_p$
1	1	4.539	0.739	1	0.8	1	1
2	7	4.112	0.727	0.830	0.8	0.88	0.875
3	3	3.760	0.908	0.665	0.8	0.96	0.428
4	17	3.586	0.570	0.977	0.7	0.68	0.659
5	5	3.420	0.818	0.466	0.9	0.92	0.316
6	29	3.351	0.863	0.947	0.4	0.44	0.701
7	2	3.350	0.528	0.663	1	0.98	0.180
8	4	3.260	0.496	0.689	0.9	0.94	0.235
9	13	3.181	0.823	0.590	0.8	0.76	0.209
10	15	3.176	0.837	0.482	0.7	0.72	0.437
13	10	3.096	0.694	0.654	0.7	0.82	0.228
22	8	2.582	0.547	0.209	0.9	0.86	0.067
23	9	2.556	0.531	0.399	0.5	0.84	0.287
24	6	2.426	0.552	0.175	0.7	0.90	0.099

表 2 で示す「センター」での結果だが、ニュース検索日 1 月 17 日の前週末に行われたセンター試験の内容を意図して検索を行った。結果上位 10 件のうち、センター試験の内容を述べている記事は 7 件であり、リランキング後の 5 位、9 位、10 位

はそれぞれ労働センターや救急センターなど、センター試験とは関係がないユーザの検索意図には合わない結果となった。

ここで注目したいのは、福島県のセンター試験受験生を扱った 6 位の記事である。まず十分なテキストであるかを示す単語量だが、上位 10 件のうち 2 番目に高い。さらに最も高い 3 位の記事はソーステキスト抽出の際にリンクのアンカテキストを取得していたことがわかったため、実質 1 番高いと言える。また類似度も 3 番目に高い値を示し、クエリ TextRank も 2 番目に高い。結果として検索エンジンにおいて 29 位という下位から 20 件以上上昇し、取得できたことは、有識性による評価が効いた一つの結果であると言える。

次に検索意図と合わない記事に注目する。まず 5 位の記事だが、時間的側面と元検索順位はその日の記事であり、検索エンジンでも 5 位のため高い値を示している。続いて、より検索意図の合致を示すと思われる類似度とクエリ TextRank を見る。まず類似度だが、上位 10 件では最低の値を示している。さらに類似度で見た場合、最下位の 3 件は 5 位、10 位、9 位であり、全てセンター試験とは関係のない記事だった。このことから、「センター」という検索クエリに関しては類似度によって検索意図を反映することができると思われる。

クエリ TextRank を見ると、上位 10 件では 4 番目の値だった。クエリ TextRank の場合、最も低い値だったのは 7 位の記事で、9 位、10 位の記事はそれぞれ 2 番目と 5 番目だった。最も低かった 7 位の記事は、センター試験の自己採点を行うという記事だった。TextRank の算出方法から考えると、クエリワードと隣接する単語が多く、さらにその単語とクエリのペアが頻出し、しかしその単語は他の単語とはあまり隣接しないという場合に TextRank は高いポイントになる。つまりクエリワードがテキスト中に頻出し、様々な単語と近接するか、同様の単語と何度もペアとなることで TextRank は上昇する。7 位の記事は 2 回しか「センター」という単語が現れていない。クエリワードが頻出することで、そのワードがニュースの主題であると言えるが、クエリ TextRank の値だけから検索意図にそっているかどうかは今回の結果からは言うことが出来ない。だが、クエリワードに注目することが間違いではないと思われる。その前後に隣接する単語を調べ、定量化することで、検索意図にあったニュースかどうかを言えるのではないかと考えられる。例えば今回の場合、「センター」というクエリワードに対して、「救急」や「処理」といった単語は意図に合わないが、「試験」という単語があれば意図に合う。このようなクエリワードに近接する単語の頻度や傾向を数値化することは有用ではないと思われる。

またリランキングで上位 10 件よりも、下位に順位を下げてしまった元検索順位 6, 8, 9, 10 位の 4 件中、8, 9 位の 2 件のニュースはセンター試験とは関係のない記事で、6, 10 位はセンター試験に関係のある記事だった。8, 9 位となった記事の類似度は、上位 10 件で最も低かった値よりもさらに低い値を示した。この結果からも類似度は検索意図を示す効果があるといえる。

検索意図に合致するにも関わらず、検索順位が下がってし

まった元 6, 10 位はそれぞれ, センター試験中に試験官の携帯がなったという記事と, 韓国語の試験に韓国の歌手グループである, 少女時代が出たという記事だった. 元 10 位の記事は有識性の値の面では現 10 位の記事と約 0.1 差とほとんど変わらず, 検索意図を示す類似度ではむしろ高い値を示している. 単語量とクエリ TextRank の値が低いため, 上記したようにクエリワード面での改善が必要かと思われる.

表 3 「バルサ」における結果

現順位	元順位	有識性	$word_p$	$sim_p$	$time_p$	$rank_p$	$que_p$
1	1	4.021	0.548	0.732	1	1	0.741
2	6	3.939	0.690	1	1	0.90	0.349
3	24	3.426	0.604	0.882	0.4	0.54	1
4	7	3.346	0.603	0.950	0.6	0.88	0.313
5	16	3.219	0.781	0.963	0.5	0.70	0.274
6	9	3.200	0.635	0.842	0.4	0.84	0.483
7	3	3.186	0.948	0.414	0.6	0.96	0.263
8	5	3.132	0.941	0.410	0.6	0.92	0.261
9	25	3.033	0.639	0.918	0.3	0.52	0.655
15	8	2.964	0.572	0.700	0.7	0.86	0.131
24	2	2.609	0.518	0.111	1	0.98	0
28	4	2.519	0.478	0.275	0.6	0.94	0.126
30	10	2.458	0.461	0.688	0.4	0.82	0.088
41	50	1.919	0.531	0.313	1	0.02	0.055

表 3 で示す「バルサ」の場合だが, クラブワールド杯の決勝でバルセロナが勝利をおさめ, 優勝したことに関連する内容を期待した. 結果上位 9 件中 8 件がバルセロナについての記事だった. 3 位となった記事はバルセロナの試合を観た, なでしこジャパンの川澄選手が, 試合で手本としたいというコメントを報じた記事で, 直接バルセロナに関係する記事ではなかった. またバルセロナに関連した記事でも, クラブワールド杯優勝に関連した記事は 8 件中 4 件しかなく, 他の 4 件は決勝前の内容, さらに準決勝前の記事という時間的に全くタイムリーでない古いものも存在した.

元検索順位がそれぞれ 2 位, 4 位, 8 位, 10 位だった記事が, 上位 10 件から外れた. そのうち 15 位, 24 位, 30 位となった 3 件の記事はクラブワールド杯決勝戦とは関係のない記事だった. 28 位となった記事は対戦相手のサントスで注目されていた選手を抑えた選手へのインタビュー記事で検索意図に合致するものだったが, 20 位以上検索順位をおとしてしまった. 170 文字程度と記事のテキスト量が少なく, 単語量も上位 10 件と比べて低かったこともあるが, 類似度, クエリ TextRank が低かったことが順位をおとした原因であると思われる. 記事中で「バルサ」という単語はタイトルに一度しか現れず, 類似度の計算対象となる単語も 5 件と少なく, さらにそれらと単語がマッチする記事は 50 件中 5 件から 7 件程度と少なかった. このことから大枠では同様の内容, 今回ならばクラブワールド杯決勝のことを報じていても, 各選手をクローズアップした記事であると, 類似度が下がってしまうことがわかった.

今回の「バルサ」の検索では, 決勝戦翌日の 12 月 19 日にニュース検索を行い, 決勝戦, もしくは優勝に関する記事を意

図したが, その意図に時間的に合った記事は上位 9 件中 4 件しかなかった. 1 位, 4 位, 7 位, 8 位がその記事だったが, ここで時間的側面のポイント  $time_p$  を見る. それぞれ 1, 0.6, 0.6, 0.6 のポイントが割り振られている. 時間的側面の項で説明したが, 0.6 は検索時刻と記事の最終更新時刻の差が, 8 時間以上 16 時間未満の場合のポイントである. ここで正確な時刻を記すと, クラブワールド杯決勝戦の終了時刻は 12 月 18 日の 21 時 30 分, そして今検索を行った時刻は 12 月 19 日の午前 10 時 8 分である. その時間差は約 12 時間半であり, 16 時間未満の時間差となる. つまりポイントが 0.6 以上ならば, その記事は今検索においては, 少なくとも時間的には価値があるタイムリーな記事ということができる. では上位 9 件中最も  $time_p$  が低かった記事は 9 位の記事でポイントは 0.3 である. この記事はワールド杯準決勝前のバルセロナの練習を報じた記事であり, 今回の検索意図には合っていない. また, 0.3 は 64 時間以上 128 時間未満の時間差, 大体 2 日以上 5 日未満の時間が経過した記事となる. 今回提案したシステム上ではポイント差として 0.6 と 0.3 で 0.3 の差しかついていないが, 「バルサ」の検索においては  $time_p$  が 0.6 未満の記事は著しく価値を落としていると言える.

「バルサ」での実験によって, 検索で意図する内容によっては時間的側面の価値が決まる境界線が異なることがわかった. 記事が増えるタイミング等の要素を考え, その話題に応じた, 時間のボーダラインを定めることが重要であると考えられる.

表 4 「箱根」における結果

現順位	元順位	有識性	$word_p$	$sim_p$	$time_p$	$rank_p$	$que_p$
1	2	3.740	0.701	0.716	0.8	0.98	0.544
2	1	3.567	0.729	0.821	0.6	1	0.417
3	25	3.441	0.570	0.851	0.5	0.52	1
4	14	3.358	0.624	0.843	0.5	0.74	0.650
5	32	3.233	0.600	0.770	0.9	0.38	0.583
6	8	3.198	0.576	0.592	0.7	0.86	0.469
7	23	3.137	0.931	0.674	0.4	0.56	0.571
8	6	3.123	0.452	0.670	1	0.90	0.100
9	3	3.081	0.843	0.554	0.4	0.96	0.324
10	28	3.030	0.619	0.792	0.5	0.46	0.659
13	5	2.900	0.749	0.355	0.8	0.92	0.077
28	4	2.559	0.447	0.196	0.8	0.94	0.176
36	9	2.272	0.665	0.489	0	0.84	0.278
39	15	2.148	0.591	0.137	0.7	0.72	0
42	10	2.063	0.638	0.469	0	0.82	0.136
48	18	1.687	0.538	0.240	0	0.66	0.249

表 4 で示す「箱根」の場合だが, 年始に行われた箱根駅伝に関連する記事を意図した. 結果として上位 10 件中 9 件が箱根駅伝に関する内容の記事だった. 9 位になった記事のみ, 東洋大が総合優勝したことにより, 東洋大の受験者増を期待するという記事だった. 検索エンジン時には検索順位 3 位であり, 順位を落としているが, 意図した内容ではないため, より順位を下げていて欲しい. 上位 10 件中類似度は最も低く, クエリ TextRank は 2 番目に低かった.

上位 10 件から外れた記事は、箱根温泉のキャンペーンに関するものや、箱根駅伝で走った選手のテレビ出演時、または twitter でのコメント等を扱った記事で、直接関係のあるものはなかった。しかし、検索エンジンで 9 位だった記事がリランキング後は 36 位となっていた。内容は一人の選手に注目したもので、二年連続で区間賞をとったという記事だった。「バルサ」の時にも述べたように、ある事柄に注目した記事は、用いる単語が他の記事と異なるため、類似度とクエリ TextRank があがりにくいことから、この記事も上位 10 件と比べると低い値である。また時間的側面の値が 0 であるが、これは時間表現を成型する際に、西暦の設定を年末の状態のままシステムを実行してしまったためのミスであった。しかし、このミスから、時間的側面のポイント  $time_p$  は、他の要素に比べ高いポイントを出しやすいため、このように取得ミスをすると大きい影響を与えることがわかった。

6 位に入った記事は、テレビ中継の視聴率に加え、優勝した東洋大の選手へのインタビューもあり、よい記事だったと言える。しかし、ソースとして取得した場合はそうだったが、実際に、リンクでニュース記事のページへいくと、ポップアップの広告ページが立ち上がり、クリックしなければ記事を読めないページだった。今回提案したシステムでは、このような場合の評価は含まれていないが、実際にユーザがページを閲覧して、このようなページを開いた場合、不快に感じるのではないかと思われる。このことから、ポップアップ広告の有無や、閲覧時に見やすいかどうかのページデザイン等の使いやすさなど、ユーザにより近い評価も考えることが重要である。

### 4.3 考察

例示した結果について考察していく。まず今回の実験では内容量があり、検索結果の他記事との類似度、クエリとの関連が高い、さらにタイムリーで検索エンジンでも評価が高い記事が、有識性が高い記事であると仮定している。ニュース検索時にクエリから連想される話題、つまり今回の実験ではユーザの検索意図に合致するであろう記事が有識性が高くなり、上位 10 件以内で取得できると想定した。

今回、「センター」「バルサ」「箱根」の 3 つのクエリについてリランキングの実験を行った。まず各クエリにおいてリランキング後に、検索意図にあう結果が上位 10 件中どの程度得られたかを見る。表 5 にまとめて示す。

表 5 上位 10 件ニュース取得結果

クエリ	検索エンジン	リランキング
センター	7	7(2 外 2 入)
バルサ	5	5(3 外 3 入)
箱根	6	9(3 外)

リランキングの括弧内の数字は、検索意図に合致しない記事が上位 10 件からいくつ外れいくつ入ったかを示している。例えば (2 外 2 入) なら 2 件、意図と合わない記事が上位 10 件から外れ、2 件の意図に合わない記事が、上位 10 件以内に入ったことを表している。

表 5 では、リランキング前後で記事数では変化なし、もしくは

は関連する記事が若干増え、よくなったという結果が得られた。また各検索において、上位に意図と合致するニュース記事がランキングされる以外にも、意図と合致しない記事、もしくは既に消去され読むことが出来ない記事の順位を下げるという結果も得られている。例えば「箱根」の検索において 18 位から 48 位となった記事、15 位から 39 位となった記事はそれぞれ箱根駅伝中に流れたビールの CM に関する記事や、駅伝参加校の OB に関する記事で意図とは関係ない記事であった。

類似度  $sim_p$  の検索意図との合致は、各実験結果で得られた。リランキング後の上位 10 件を見た場合、意図と関係が薄い記事の類似度は最も低い値を示している。ただし、この低いというのは相対的に見てであって、値としてみた際には、0.7 を超えるような高い値となる場合もあった。

次いで、単語量  $word_p$  についてである。 $word_p$  を用いることで、ニュース記事において有益なテキストの量を判断できると考えたが、今回の実験において、 $word_p$  の値は他要素に比べて、ニュース記事ごとの差が出にくいものとなった。これは計算方法の違いからくるものであると思われる。 $sim_p$  と  $que_p$  に関しては、各ニュース記事における計算で得られたコサイン類似度の平均や、クエリワードの TextRank を、その検索における最大値で割って値を求めている。しかし  $word_p$  に関しては、非重複単語数を検索における最大非重複単語数で割った値と、各記事の全単語数で割った値の和の半分の値が  $word_p$  となる。非重複単語数を全単語数で割った値は低い値にはなりにくく、結果として差が出にくい。しかし、 $word_p$  においてそれぞれの値は、検索におけるテキストの量を示し、また各記事における有益なテキストの量を示す上で必要だと考えられる。よって、 $word_p$  でより記事間の差を出すための計算法の改善、または正規化を行う必要がある。

次に、各話題における句の時間の存在について考える。「バルサ」でも述べた通り、各検索クエリにおいて流行する話題には、ニュース記事に価値のある時間帯、句の時間が存在する。本研究で用いている  $time_p$  の時間区分は、単純に時間を倍々にしていったものを使用した。しかし、価値のある時間帯はクエリごとに異なることがわかったため、クエリごとに適した時間区分の手法を考えることが必要である。例えば、検索エンジンにおいてクエリとして選択した単語を含んだタイトル数が目立って増加した時間を調べることができれば、その時間帯を境として、より適した  $time_p$  のポイント振り分けが可能となるかもしれない。

### 4.4 テキスト取得改善実験

#### 4.4.1 実験準備

前実験においては、Web ニュースのテキスト取得時に本文以外のアンカテキスト等も合わせて取得していることがわかった。これらの余分なテキストは類似度や単語量、クエリ TextRank の計算に影響し、有識性を誤った値とする。その為、リランキング結果を不正確にする可能性がある。本実験ではニュース本文の取得ミスの改善による効果を検証するため、ニュース記事の本文取得を手動で行った場合と比較する。この実験を行うことで、本文からアンカテキストのような、ニュースには関係の

ないテキストを除くことができ、内容に関係なく有識性の値が変化することを無くすことができると考えられる。

手動の場合は、システムへの本文入力手順を新たに設け、そこへ変数として本文を入力する。今回本文として取得するのはニュースタイトルとニュース記事の本文である。一つの文として入力するため、タイトルの末尾には『。』を付け、本文とつなげて入力する。クエリには「大阪」を用いた。検索したのは2011年12月20日で、橋下市長が当選した翌日である。

表6に「大阪」での検索結果の一部を示す。自動としているのは前実験と同手法で計算した際の値であり、手動として括弧中に示している値は本文を手動で入力して得られた値である。なお  $rank_p$ ,  $time_p$  は自動、手動問わず同じ値が出るため割愛する。

表6 「大阪」における結果

手動順位 (自動順位)	有識性	$word_p$	$sim_p$	$que_p$
1 (1)	4.031 (3.869)	0.748 (0.846)	0.979 (0.760)	0.607 (0.565)
2 (2)	3.813 (3.571)	0.885 (0.571)	0.927 (1)	1 (1)
3 (4)	3.755 (3.414)	0.925 (0.709)	0.594 (0.610)	0.664 (0.548)
4 (6)	3.474 (3.268)	0.512 (0.566)	0.939 (0.690)	0.544 (0.532)
5 (3)	3.226 (3.445)	0.588 (0.791)	0.777 (0.811)	0.345 (0.327)
6 (9)	3.163 (2.929)	0.829 (0.578)	0.602 (0.677)	0.293 (0.235)
7 (5)	3.038 (3.352)	0.396 (0.562)	0.516 (0.604)	0.168 (0.227)
8 (12)	2.876 (2.748)	0.597 (0.610)	0.925 (0.719)	0.443 (0.530)
9 (7)	2.860 (3.182)	0.595 (0.788)	0.620 (0.758)	0.308 (0.299)
10 (11)	2.850 (2.822)	0.748 (0.760)	0.402 (0.465)	0.322 (0.219)
41 (16)	1.759 (2.571)	0.441 (0.798)	0.435 (0.784)	0.160 (0.287)

#### 4.4.2 考察

本文テキストを手動で取得した場合が、システムで自動的に取得した場合よりも、 $word_p$  の値を下げている記事は上位10件中7件、 $sim_p$  も7件、 $que_p$  は2件だった。取得するテキストが減るため、 $word_p$  は全ての記事が減ると考えていたが、上昇した記事も存在した。上位10件中上昇したのは3件で、2位、3位、6位の記事だった。これは取得できる単語数は減るが、同時に重複によって除かれる単語も減るため、値が高くなったと考えられる。 $sim_p$  が上昇した記事は1位、4位、8位の記事だが、これも単語が減ったことで、全ての記事において単語のTextRankが低下し、その中でもTextRankの低下幅が低いもの、また単語数の減少が少ない記事がポイントをあげたのではないかと思われる。 $que_p$  にも同様にクエリワードのTextRank

の値低下による影響であることが言える。

取得精度の上昇はニュース提供しているサイトによって効果が変わるが、アンカテキストが大量に混ざった結果、本文テキストは少ないが有識性の値が高くなり上位となるようなミスは精度向上により改善できると思われる。ただし、本来の目的であるよりよい記事の取得が、本文の取得精度の改善のみで達成できるとは「大阪」というクエリにおいては言い難い。今回のクエリでの自動と手動の検索順位変動平均は4.85だった。取得精度を高めることは重要だが、5件弱の変動が本文取得ミスから起こるとしても、よりよい記事の取得が失敗した際の全ての理由がここにあるとは言えない。

## 5. おわりに

今回、ユーザの検索意図にあった有益なオンラインニュース取得のため、ニュース記事から要素を抽出し計算する有識性を提案し、その有識性によるオンラインニュースのリランキング実験と、その考察を行った。実験の結果から有識性を用いてオンラインニュースをリランキングすることで、検索に不慣れたユーザを想定した場合、その検索意図を反映した有益なニュース記事取得結果を得ることが出来た。特に、有識性構成要素である類似度  $sim_p$  においては、よい結果を得ることができ、今後の指標としていきたい。同時に、改善すべき問題の存在という知見も得ることができた。単語量  $word_p$  の計算法や、クエリワードのみでなく、それに近接する単語を考慮することで、よりよい結果が得られると思われる。

今後の課題としては、提案した有識性の有効性向上のために、有識性そのものの定義をさらに検討する必要があると考えている。また、システムの高速度化、本文や更新時刻といった要素取得精度の向上、各ニュースごとのタイムリーな時間の推測、適切な重みづけの改善も必要である。

## 文献

- [1] Shawn Tseng and B.J. Fogg: Credibility and Computing Technology, CACM, Vol.42, No.5, pp. 39-42 (1999).
- [2] B.J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani and Marissa Treinen: What Makes Web Sites Credible?: A Report on a Large Quantitative Study, SIGCHI2001, pp. 61-68 (2001).
- [3] 加藤義清, 黒橋禎夫, 江本浩: 情報コンテンツの信頼性とその評価技術, 人工知能学会 SIG-SWO, A602-01 (2006).
- [4] Soo Young Rieh, Yong-Mi Kim, Ji Yeon Yang and Beth St. Jean: A Diary Study of Credibility Assessment in Everyday Life Information Activities on the Web: Preliminary Findings, ASIST 2010, Vol.47, No.1, pp. 1-10 (2010).
- [5] Alia Amin, Junte Zhang, Henriette Cramer, Lynda Hardman and Vanessa Evers: The Effects of Source Credibility Ratings in a Cultural Heritage Information Aggregator, WICOW'09, pp. 35-42 (2009).
- [6] Ryosuke Nagura, Yohei Seki, Noriko Kando and Masaki Aono: A Method of Rating the Credibility of News Documents on the Web, SIGIR'06, pp. 683-684 (2009).
- [7] Rada Mihalcea and Paul Tarau: TextRank: Bringing Order into Texts, In Conference on Empirical Methods in Natural Language Processing (2004).