

Wikipedia を知識源とするニュース・ブログ間の相補的ナビゲーション

佐藤 由紀[†] 中崎 寛之^{††} 川場真理子^{††} 宇津呂武仁^{††} 福原 知宏^{†††}

[†] 筑波大学第三学群工学システム学類 〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

^{†††} 東京大学 人工物工学研究センター 〒 277-8568 千葉県柏市柏の葉 5-1-5

あらまし 本研究では、検索エンジン等を用いた検索行動のうちでも、特に、客観的かつ恒久的な事実を記載した Wikipedia、詳細な事実情報を報道するニュース、および、個人の主観的意見や経験などを豊富に記載したブログの検索に焦点を当てて、利用者の検索行動を支援する枠組みを提供することを目的とする。本研究では、これらの三種類の情報源の間で、密接に関連する項目や記述部分の間を相互にナビゲートする機能を実現し、利用者の検索行動を支援する。

キーワード 情報検索, Wikipedia, ニュース, ブログ, トピック分析

Complementary Navigation of News and Blogs with Wikipedia as Fundamental Knowledge Source

Yuki SATO[†], Hiroyuki NAKASAKI^{††}, Mariko KAWABA^{††}, Takehito UTSURO^{††}, and Tomohiro
FUKUHARA^{†††}

[†] College of Eng. Sys., Third Cluster of Colleges, University of Tsukuba, Tsukuba, 305-8573, Japan

^{††} Grad. Sch. Systems and Information Engineering, University of Tsukuba, Tsukuba, 305-8573, Japan

^{†††} Research into Artifacts, Center for Engineering, University of Tokyo Kashiwa, Chiba 277-8568, Japan

Abstract This paper studies complementary navigation of news and blog, where *Wikipedia* entries are utilized as fundamental knowledge source for linking news articles and blog feeds/posts. In the proposed framework, given a topic as the title of an *Wikipedia* entry, its *Wikipedia* entry body text is analyzed as fundamental knowledge source for the given topic, and terms strongly related to the given topic are extracted. Those terms are then used for ranking news articles and blog feeds/posts. In complementary navigation of news and blog, *Wikipedia* entries are retrieved when seeking fundamental background information, while news articles are retrieved when seeking precise news reports on facts, and blog feeds/posts are retrieved when seeking subjective information such as personal opinions and experiences.

Key words IR, Wikipedia, news, blog, topic analysis

1. はじめに

本研究では、検索エンジン等を用いた検索行動のうちでも、特に、客観的かつ恒久的な事実を記載した Wikipedia、詳細な事実情報を報道するニュース、および、個人の主観的意見や経験などを豊富に記載したブログの検索に焦点を当てて、利用者の検索行動を支援する枠組みを提供することを目的とする。本研究では、これらの三種類の情報源の間で、密接に関連する項目や記述部分の間を相互にナビゲートする機能を実現し、利用者の検索行動を支援する (図 1)。

Wikipedia, ニュース, ブログの三者を比較すると, Wikipedia

は、インターネット上の最大規模の百科事典として、近年、様々な研究分野において利用されている (例えば, [1], [2])。日本語では、約 55 万のエントリ (2008 年 1 月現在) が収録されており、しかも、多くの人が自由にエントリを書くことができるため、ニュースやブログで話題となる事項のエントリが、迅速に作成されるという特徴を持っている。Wikipedia を利用した研究事例としては、図書館の分類体系と Wikipedia カテゴリの対応付けを行う研究 [1] や、Wikipedia の言語間リンクを利用して多言語対訳辞書を作成するという研究 [2] などがある。

ニュースとブログを比較すると、ニュースは、従来より、日々の報道を閲覧するという形で利用されてきた。一方、ブログに

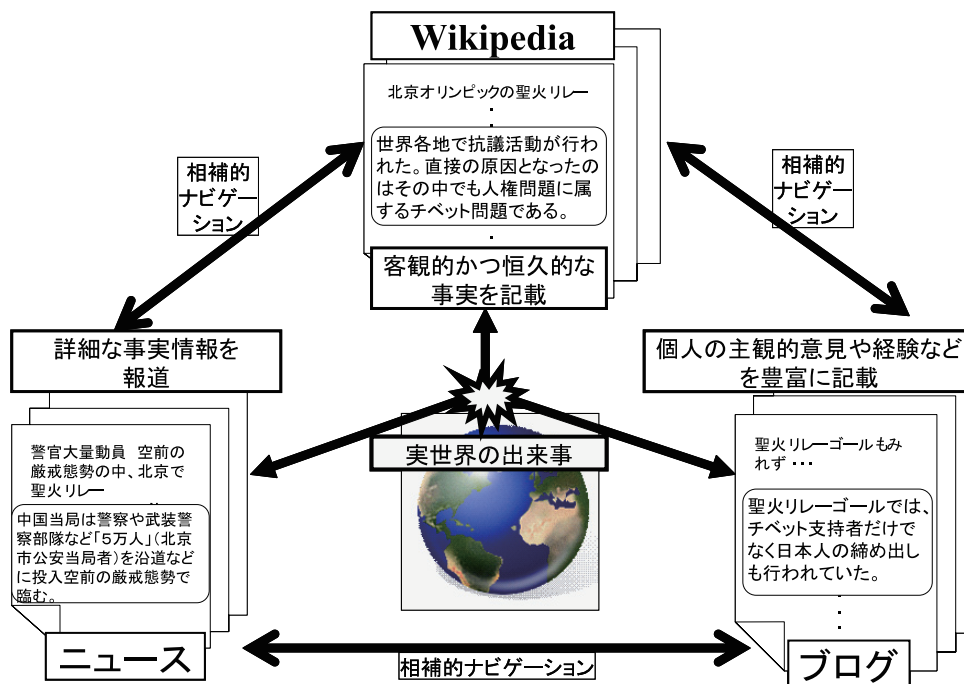


図1 Wikipedia, ニュース, ブログ間の相補的ナビゲーションの枠組み

ついても近年、世界中でブログサービスやブログツールが普及し、各地域の人々がそれぞれインターネット上で個人の意見や評判を発信することが可能になるのに伴って、様々な情報がブログに記載され、また、商用ブログ検索サービスを利用することでそれらの情報を取得することができるようになった。具体的なサービスの例として、*Technorati*^(注1)、*BlogPulse*^(注2)、*kizasi.jp*^(注3)、*blogWatcher*^(注4)などが挙げられる。これらの検索サービスは、巨大なブログ空間の索引付けという観点から見ると、キーワードや評判、時系列変化や人手によって作成されたカテゴリ情報などを索引として用いて、利用者の求めるブログ記事やブログサイトを検索する。

ここで、本論文において、Wikipedia、ニュース、ブログの三種類の情報源の間で、密接に関連する項目や記述部分を相互にナビゲートする機能を実現するにあたっては、まず、あるトピックについて、Wikipediaのエントリから関連する用語を抽出し、これらの用語を知識源として、ニュース、ブログから関連するニュース記事、ブログサイト、ブログ記事を検索する。この検索のうち、特にブログサイトおよびブログ記事の検索においては、我々はすでに、[3]~[5]において、Wikipediaエントリの記述内容をトピックとする有用なブログサイトおよびブログ記事を検索する方式を確立している。この方式においては、Wikipediaエントリ名を表すキーワードを用いて商用検索エンジンAPIにより上位のブログサイトを収集し、これを、当該キーワード、およびWikipediaエントリから抽出した関連語の出現数順に順位付けするという要素技術を用いている。

一方、本論文では、同様の検索手法をニュース記事の順位付けにも用いることにより、Wikipediaエントリの記述内容をトピックとするニュース記事を選別する方式の有効性を示す。さらに、ニュース記事の順位付けにおいては、Wikipediaエントリから関連語を抽出するだけでなく、エントリ本文テキストから名詞句を抽出し、この名詞句と関連語を併用することにより、ニュース記事を順位付けする方式の有効性を示す^(注5)。

また、上記の、Wikipediaを知識源としてニュース、ブログから関連する項目や記述部分を検索する方式を一般化すると、Wikipediaを知識源としなくても、ニュース、ブログ間で関連する項目や記述を相補的に検索する、あるいは、ニュース、ブログを情報源として、関連するWikipediaエントリを検索する、という方向でのナビゲーションの実現が可能となる。本論文の研究においては、今後、そのような柔軟な方向性を持った、Wikipedia、ニュース、ブログ間の相補的ナビゲーションの研究を進める。

2. Wikipedia エントリからの関連語抽出

ニュース記事およびブログ記事の検索において、Wikipediaエントリを知識源として用いるために、エントリ本文から当該トピックの関連語を抽出する。本論文においては、当該エントリのリダイレクトのタイトル、エントリ本文中の太字、エントリ本文中においてリンクされている他エントリのタイトル、本

(注1) : <http://technorati.com/>

(注2) : <http://www.blogpulse.com/>

(注3) : <http://kizasi.jp/>(日本語のみ)

(注4) : <http://blogwatcher.pi.titech.ac.jp/>(日本語のみ)

(注5) : 本論文の執筆段階においては、開発途中のため、ニュース記事の順位付けには全ての種類の用語を用いているが、ブログサイト・ブログ記事の順位付けにおいては、Wikipediaエントリのタイトル、関連語のみを用いている。また、今後、ニュース記事、および、ブログサイト・ブログ記事の順位付けにおいて、Wikipediaエントリのタイトル、関連語(リダイレクトおよび強調文字、リンク)、エントリ本文テキスト中の名詞句等の数種類の用語について、個別に効果の評価を行う予定である。

文中の各段落のタイトル, および, 本文テキスト中の全名詞句を関連語として抽出する [3]~[5].

3. Wikipedia エントリからのニュース記事・ブログ記事の検索

3.1 ニュース記事検索

Wikipedia エントリをトピックとするニュース記事の検索においては, Wikipedia エントリのエントリ名を検索クエリとして, 検索クエリを含む記事を全て収集した.

3.2 ブログ記事検索

3.2.1 ブログサイトの収集

Wikipedia エントリをトピックとするブログサイトの収集においては, Yahoo!Japan 検索 API を利用し, 大手 11 社^(注6)のブログホストに限って検索を行った. 検索の際には, Wikipedia エントリのエントリ名を検索クエリとして, 複数のブログホストを一度に指定して検索し, 1000 件の記事を取得する. しかし API の検索ではブログ記事単位の検索になるので, 同一著者のブログ記事は一つのブログサイトにまとめるという作業を行った. その結果, 一トピックあたり約 200 前後のブログサイトを取得することができた. その後, 各ブログサイトにおいて, Wikipedia エントリのエントリ名のヒット数を求め, ヒット数が下限未満 (本論文では, 10) のブログサイトを削除した.

3.2.2 ブログ記事の選別

次に, 収集されたブログサイト中のブログ記事のうち, 検索トピックに関連のある記事のみを選別するために, 2. 節の手順により Wikipedia エントリから抽出した関連語が出現する記事のみを選別する. 具体的には, 当該 Wikipedia エントリのリダイレクトのタイトル, エントリ本文中の太字, および, エントリ本文中においてリンクされている他エントリのタイトルを関連語として抽出し, それらの関連語のいずれかが出現する記事のみを選別する.

3.3 Wikipedia エントリとニュース記事・ブログ記事の間の類似度

検索されたニュース記事およびブログ記事の Wikipedia エントリとの類似度算出においては, 2. 節の手順により Wikipedia エントリから抽出した関連語を用いる. 具体的には, 2. 節において抽出された関連語 t の種類 $type(t)$ ごとに重み $w(type(t))$ を決めておき, 以下の総和によって, Wikipedia エントリ E およびニュース記事・ブログ記事 D の間の類似度 $Sim_{w,nb}(E, D)$ を定義する.

$$Sim_{w,nb}(E, D) = \sum_t w(type(t)) \times freq(t)$$

ただし, $freq(t)$ は, 記事中における関連語 t の出現頻度である.

ここで, 関連語 t の種類 $type(t)$ ごとの重み $w(type(t))$ は, ニュース記事の類似度算出においては,

$$w(\text{リダイレクト}) = w(\text{太字}) = w(\text{段落タイトル}) =$$

$$w(\text{本文名詞句}) = 1$$

$$w(\text{他エントリ・リンク}) = 0$$

とし, ブログ記事の類似度算出においては,

$$w(\text{リダイレクト}) = 3, w(\text{太字}) = 2$$

$$w(\text{他エントリ・リンク}) = 0.5$$

$$w(\text{段落タイトル}) = w(\text{本文名詞句}) = 0$$

とする.

3.4 ニュース記事・ブログ記事の順位付け

ニュース記事・ブログ記事の順位付けにおいては, 前節で述べた類似度の降順に記事を順位付けする.

4. ニュース記事・ブログ記事からの Wikipedia エントリの検索

ニュース記事・ブログ記事 D からの Wikipedia エントリの検索においては, ニュース記事・ブログ記事中に出現した Wikipedia エントリ名を E_1, \dots, E_n として, 3.3 節で定義した類似度 $Sim_{w,nb}(E_i, D)$ ($i=1, \dots, n$) の降順に E_1, \dots, E_n を順位付けする.

5. Wikipedia エントリを介したニュース記事とブログ記事の対応付け

ニュース記事もしくはブログ記事を検索質問として, 4. 節の手順によって検索結果として得られる Wikipedia エントリを入力として, 3. 節の手順によって, 関連するブログ記事もしくはニュース記事を検索することにより, 知識源として Wikipedia エントリを介することにより, トピックの関連するニュース記事・ブログ記事を対応付けることができる.

ここで, 検索質問となるニュース記事もしくはブログ記事を D_1 , 検索対象となるブログ記事もしくはニュース記事を D_2 として, 両者の間の類似度を以下の式 $Sim_{n,w,b}(D_1, D_2)$ で定義する.

$$\begin{aligned} Sim_{n,w,b}(D_1, D_2) &= Sim_{nb}(D_1, D_2) \\ &+ K_{w,nb} \sum_E \left(Sim_{w,nb}(E, D_1) + Sim_{w,nb}(E, D_2) \right) \end{aligned}$$

ただし, 両者の間に介在する Wikipedia エントリを E とする. この類似度 $Sim_{n,w,b}(D_1, D_2)$ においては, $Sim_{n,w,b}(D_1, D_2)$ は, D_1 と D_2 の間の直接の文書間類似度 $Sim_{nb}(D_1, D_2)$, および, Wikipedia エントリ E を介する際の類似度 $Sim_{w,nb}(E, D_1)$, $Sim_{w,nb}(E, D_2)$ の総和の重み付き和として定義される. D_1 と D_2 の間の直接の文書間類似度 $Sim_{nb}(D_1, D_2)$ としては, 余弦類似度や文書ベクトル間の内積等を用いる. ここで, 重み $K_{w,nb}$ の値を調整することにより, 通常用いられる文書間類似度 $Sim_{nb}(D_1, D_2)$ と, Wikipedia エントリを介した類似度との間の比率を調整する.

6. サンプルトピック・ニュース記事・ブログ記事

本論文において分析対象としたトピックの一覧を表 1 に示す.

(注6): FC2.com, yahoo.co.jp, rakuten.ne.jp, ameblo.jp, goo.ne.jp, live-door.jp, Seesaa.net, jugem.jp, yaplog.jp, webry.info.jp, hatena.ne.jp

表 1 サンプルトピックおよびサブトピック一覧

トピック	サブトピック
京都議定書	議決内容, 締約状況, 京都メカニズム, 日本の削減量の内訳と現状, 京都議定書に関する議論
ミサイル	ミサイルの種類, 新たな技術, イージス艦衝突事故, 技術漏洩, 安全保障, ゲーム (ゲーム内の武器として), MD, テポドン
ヒラリー・クリントン	大統領選以前, スーパーチューズデイ, 女性大統領誕生への期待, ヒラリー凋落, 選挙資金, 対日本のヒラリーのスタンス, 大統領選以降
バンク・オブ・アメリカ	合併前のバンクオブアメリカ社の歩み, 合併後のバンクオブアメリカ社の歩み, 株価レポート, 米金融不安, 世界金融市場と日中金融機関, メリルリンチとの合併, 米国株式市場, 韓国金融市場

表 2 Wikipedia エントリから抽出した用語数

トピック	リダイレクト	太字	他エントリ・リンク	段落タイトル	本文名詞句
京都議定書	0	14	404	9	720
ミサイル	0	1	149	14	988
ヒラリー・クリントン	0	3	247	11	622
バンク・オブ・アメリカ	0	8	81	4	333

また, 本論文において検索の対象としたニュース記事について新聞社ごとの記事数 (ニュース記事の収集時期は, 2008 年 1 月 1 日 ~ 9 月 29 日^(注7)) を表 3 に示す. さらに, 2. 節の手順により Wikipedia エントリから抽出した用語 (表 2 に数を示す) を用いて, 前節で述べた手法によりニュース記事およびブログ記事の検索・順位付けを行い, 上位 20 記事ずつを収集した. 検

表 3 新聞社ごとのニュース記事数

新聞社	ニュース記事数
朝日新聞	33039
読売新聞	26657
日経新聞	39164
CNN 日本語版 (アメリカ)	5344
朝鮮日報日本語版 (韓国)	13882
中央日報日本語版 (韓国)	10488
人民網日本語版 (中国)	11697
合計	140271

表 4 検索されたニュース記事数およびブログサイト数・ブログ記事数 (各トピックごと)

トピック	ニュース記事数	ブログサイト数	ブログ記事数
京都議定書	413	97	1258
ミサイル	739	78	2158
ヒラリー・クリントン	663	68	470
バンク・オブ・アメリカ	264	38	506

(注7): ニュース記事については, すでに, 2006 年分以降を収集済みであり, 今後, この全体を対象として評価実験を行う予定である.

表 5 Wikipedia エントリからのブログ記事の検索性能の評価 (%)

トピック	トピック名		関連語
	無関係記事率	スプログ率	無関係記事率
京都議定書	10	0	0
ミサイル	0	0	0
ヒラリー・クリントン	0	45	5
バンク・オブ・アメリカ	55	20	0

表 6 情報源間のサブトピック共有率 (各トピックごと, %)

トピック	Wikipedia-ニュース間	Wikipedia-ブログ間	ニュース-ブログ間
京都議定書	75	88	82
ミサイル	68	33	73
ヒラリー・クリントン	49	39	62
バンク・オブ・アメリカ	28	6	62

索されたニュース記事数およびブログサイト数・ブログ記事数を表 4 に示す. さらに, 各トピックについて, Wikipedia の段落, および, 検索されたニュース記事およびブログ記事中の記述内容を人手で分析し, 表 1 のサブトピック一覧を作成した.

7. 評価

7.1 Wikipedia エントリからのニュース記事・ブログ記事の検索

7.1.1 ニュース記事・ブログ記事の検索精度の評価

3. 節で述べた手法の有効性を検証するために, トピック名の出現頻度の降順に順位付けしたものととの比較を行った.

ニュース記事の検索性能については, 関連語を用いた順位付け, トピック名を用いた順位付けともに, 表 1 のサブトピックと無関係な記事は, 三記事のみであり, 大きな違いは見られなかった.

ブログ記事の検索性能について, 表 5 に評価結果を示す. この結果からわかるように, 関連語を用いた順位付けでは, スプログ [6] が混入せず, 無関係記事もわずかであった. 一方, トピック名のみを用いた順位付けでは, 無関係記事・スプログが多く混入していた. この結果より, 関連語を用いた順位付けの有効性が確認できた.

7.1.2 情報源間のサブトピック共有率

ここでは, 表 1 に挙げた各トピックについて, Wikipedia エントリ中の段落, ニュース記事, ブログ記事の記述内容を人手で比較して, 表 1 中のサブトピックがどの程度共有されているかを測定する. 具体的には, まず, Wikipedia の各一段落, ニュース記事一記事, ブログ記事一記事を, それぞれ, サブトピックに関する記述の有無を次元とするベクトル d_w , d_n , d_b (ただし, 各ベクトルの大きさは, $|d_w| = |d_n| = |d_b| = 1$ に正規化する) によって表現する. 次に, 一つのトピックについて, Wikipedia エントリ中の全段落, 検索結果の全ニュース記事 (本論文では, 20 記事), および, 検索結果の全ブログ記事

表 7 ニュース記事からの Wikipedia エントリ検索における関連エントリ率 (%)

トピック	タイトル	関連語	本文名詞句
京都議定書	34	46	86
ミサイル	58	94	98
ヒラリー・クリントン	52	58	78
パンク・オブ・アメリカ	34	70	86

(本論文では 20 記事) から、各ベクトルの総和 $\sum d_w, \sum d_n, \sum d_b$ を求める。そして、これらの和ベクトルの間の余弦によって、二つの情報源間のサブトピック共有率を定義する。

$$\text{情報源 } ij \text{ 間のサブトピック共有率} = \frac{(\sum d_i) \cdot (\sum d_j)}{|\sum d_i| |\sum d_j|}$$

このようにして求めたサブトピック共有率を表 6 に示す。

7.1.3 各情報源中の記述の比較分析

表 9~12 に、各トピックのサブトピックについて、Wikipedia, ニュース, ブログに共通する記述内容、および各情報源特有の記述内容の抜粋を示す。この結果から分かるように、Wikipedia で解説されている事項のうちの多くは、ニュースもしくはブログでも取り上げられている。一部では、ニュースでは検索されなかったサブトピックも存在するが、その原因の大半は、ニュース記事の収集時期が短いため、および、調査対象としたニュース記事数が 20 件と少ないためであると考えられる。また、Wikipedia, ニュースともに、情報源特有の記述が観測されている。このうち、Wikipedia 特有の記述内容については、ニュース記事の収集時期を拡大することにより、同一事項に関する記述の有無が正確に判断できると考えている。今後、以上の分析結果をふまえて、Wikipedia, ニュース, ブログにおける個々の記述内容を検索クエリとして、関連記事検索を行う予定であり、この研究の一環において、同一事項の記述の有無の自動判定に取り組む予定である。

7.2 ニュース記事からの Wikipedia エントリの検索

4. 節で述べた手法により、ニュース記事からの Wikipedia エントリの検索の評価を行った。ただし、ニュース記事 D と Wikipedia エントリ E との間の類似度 $Sim_{w, nb}(E, D)$ (3.3 節参照) においては、Wikipedia エントリのタイトルのみを用いる場合、Wikipedia 関連語として、「リダイレクト」、「太字」、「他エントリ・リンク」のみを用い重みを全て 1 とする場合、および、本文名詞句のみを用いる場合の三通りを比較した。

各トピックについて、ニュース記事 5 件を検索質問として、検索されたエントリの上位 10 位までに、検索質問のニュース記事と関連する Wikipedia エントリが含まれる割合を表 7 に示す。この結果より、Wikipedia 本文名詞句のみを用いて類似度算出を行うことにより、ニュース記事に該当する Wikipedia エントリを正確に判断できると考えられる。また、本文名詞句のみを用いてニュース記事と Wikipedia エントリの類似度を測定した場合について、Wikipedia エントリの検索事例を表 8 に示す。この結果から分かるように、非関連エントリの多くが一

般語であるので、Wikipedia 階層構造中で上位に位置するか否か、あるいは、ウェブ検索エンジンを用いて測定したヒット数が一定値以下であるか、等の条件を課すことにより、これらの非関連エントリが排除できる可能性がある。

8. 関連研究

ニュースとブログとの間の相補的な利用については、[7]~[9] などの研究がある。[7] では、ブログ記事中で参照しているウェブサイトやニュース記事をそのユーザの興味の対象として、ブロガーの嗜好を利用したウェブ情報推薦システムを提案している。具体的にはニュースサイトとブログの対応付けを行い、ユーザの嗜好にあったニュース記事を推薦するというを行っている。ただし、ニュース記事とブログの対応付けにおいては、ブログからニュース記事への引用の有無を利用しており、ブログ記事に対するテキスト検索は行われていない。[8] では、ニュース記事とブログ記事との間の文書類似度において、語の出現頻度の推移を考慮した重み付けを用いることにより、ニュース記事に関連したブログ記事を対応付ける手法を提案している。この手法は我々の研究においても有用と考えられるので、今後、本論文の、Wikipedia エントリを知識源とする手法との併用を進める。また、[9] では、ブログ記事からニュース記事へのアンカーリンクを用いて、ニュース記事に関連するブログ記事の収集を行っている。本論文でも、ブログ記事からニュース記事へのアンカーリンクが抽出できる場合には、その情報を利用する予定であるが、アンカーリンクが抽出できないブログ記事の場合には、主として、Wikipedia エントリを知識源とする本論文の手法による対応付けを用いる。

一方、[10] では、同じ事象について、複数の情報源の情報の伝え方の異なりかたを分析することを目的として、複数の国の代表的なメディアが発信するニュースを情報源として、各々の国の世論がどのように事象を分析しているのかを把握する方式を提案している。

9. おわりに

本論文では、検索エンジン等を用いた検索行動のうちでも、特に、客観的かつ恒久的な事実を記載した Wikipedia, 詳細な事実情報を報道するニュース、および、個人の主観的意見や経験などを豊富に記載したブログの三種類の情報源の間で、密接に関連する項目や記述部分の間を相互にナビゲートする方式について提案し、三種類の情報源の記述内容の比較分析を行った結果について述べた。今後は、複数情報源間における同一記述の有無の自動判定について取り組む。

謝辞: 本論文の実験で使用したニュース記事に関して協力して頂いた、北海道大学大学院情報科学研究科 吉岡真治准教授に感謝する。

文 献

- [1] 田村悟之, 清田陽司, 増田英孝, 中川裕志. 図書館における自動レファレンスサービスシステムの実現 Web 上の二次情報と図書館の一次情報の統合. 情報処理学会研究報告, Vol. 2007, No. (2007-FI-179), pp. 1-8, 2007.
- [2] 新井嘉章, 福原知宏, 増田英孝, 中川裕志. Wikipedia を用いた

表 8 ニュース記事からの Wikipedia エントリの検索事例

トピック	検索質問として用いたニュース記事の概要	検索された Wikipedia エントリ	
		関連エントリ	非関連エントリ
京都議定書	温暖化対策に日本発の科学技術の導入	日本/地球温暖化/ヨーロッパ/環境問題/アメリカ合衆国/第 34 回主要国首脳会議/資源/インド/化石燃料	政治
ミサイル	日本の国防（北朝鮮のミサイルに対して）	軍事/自衛隊/日本/ミサイル防衛/戦争/軍隊/ノド ン/戦闘機/安全保障/弾道ミサイル	(関連エントリなし)
ヒラリー・クリントン	代表戦撤退表明	ヒラリー・クリントン / 議員 / アメリカ合衆国 / 選挙 / 副大統領 / 上院 / クリントン / 大統領	ワシントン / 委員会
バンク・オブ・アメリカ	バンク・オブ・アメリカによる、住宅ローン会社の救済合併について	サブプライムローン / バンク・オブ・アメリカ / 住宅ローン / アメリカ合衆国 / 貸金業 / 出資	住宅 / バンク / 大手 / 合意

表 9 Wikipedia, ニュース, ブログ中の記述の比較分析 (「京都議定書」)

サブトピック	共通事項	Wikipedia 特有	ニュース特有	ブログ特有
議決内容	参加各国の削減目標および、議決内容に違反した場合の処置について	(特有の記述なし)	(記事なし)	「実はこの地球上で唯一日本だけが削減義務を課せられているのです。」「民間企業に対する二酸化炭素の強制的な排出削減義務化は、多くの企業を強制倒産させ、必ず大不況を招く愚策だと思います。」
締約状況	締結から発行までの流れ、および、参加国の参加体制について	(特有の記述なし)	京都議定書に参加しないアメリカの洞爺湖サミットにおける温暖化対策関連の取り組みについて	「EU では、マイナス 6% などとちんけなことは言わずに、マイナス 20% を目標にすると、1/10 に発表したばかり。すごいですねー。えらいですねー。先進国ですねー。… 各国とも排出削減にふんばってる時に、ニホンとアメリカは逆行しちゃってるのだ。」
京都メカニズム	植林活動、国外での活動、削減量の国家間取引など、温室効果ガスの削減をより容易にするための規定について。	既存の森林を CO ₂ 吸収分として算出できるように規約を拡張した。	日本の世界最先端環境技術について、世界にアピールしていくことにより地球温暖化を防ごう。	「人間は利益がなければ地球環境を守れないということか。」「これからは、地球温暖化防止のためには、国も、企業も、ひいては国民も自分の努力で CO ₂ 削減が出来ない人はその分お金をささなくてはいけなくなるよ…」
日本の削減量の内訳と現状	日本の各分野における削減量の具体値および現状		(記事なし)	「電力会社に余剰電力を売電しても、温暖化防止にならないことはこれまで知られることがありませんでした。」
京都議定書に関する議論	日本に架せられた削減目標の重さ、京都議定書の内容での温暖化防止効果の程度、地球温暖化現象の原因、京都議定書後の取り組み、等の議論の概略	どれほどの効果があるのかという議論や、二酸化炭素の排出量を規制するのは本当に温暖化防止に役立つのかという議論の紹介	中国政府の温室効果ガス削減に関する取り組みと声明	「日本国民にとって一番の問題は、この条約で一番不利なのがどこからどうかんがえても日本だ、ということです。」

多言語ブログ検索のための訳語抽出. 情報処理学会第 70 回全国大会講演論文集, 第 5 巻, pp. 55–56. 情報処理学会, 2008.

- [3] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. Wikipedia エントリとブログサイトの対応付けのための特定トピックのブログサイト検索. DEWS2008 論文集, 2008.
- [4] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. 多言語 Wikipedia エントリを用いた特定トピックブログサイト検索と日英対照ブログ分析. 第 22 回人工知能学会全国大会論文集, 2008.
- [5] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. Wikipedia エントリとブログサイトの対応付けによる日本語ブログ空間のトピック分布推定. 情報処理学会研究報告, Vol. 2008, No. (2008-NL-187), pp. 83–90, 2008.
- [6] 佐藤有記, 宇津呂武仁, 福原知宏, 河田容英, 村上嘉陽, 中川裕志, 神門典子. キーワードの時系列特性を利用したスパムブログの取

集・類型化・データセット作成. DEWS2008 論文集, 2008.

- [7] 小原恭介, 山田剛一, 絹川博之, 中川裕志. Blogger の嗜好を利用した強調フィルタリングによる Web 情報推薦システム. 第 19 回人工知能学会全国大会発表論文集, 2005.
- [8] 池田大介, 藤木稔明, 奥村学. blog とニュース記事の自動対応付け. 言語処理学会第 11 回年次大会論文集, pp. 1030–1033, 2005.
- [9] 石崎諒, 青野雅樹. Web ニュースに対するブログ意見の分析ツール. 電子情報通信学会技術研究報告, WI2-2008-52, pp. 11–12, 2008.
- [10] 吉岡真治. 複数のニュース源の差異を考慮したニュース分析の研究. 言語処理学会第 13 回年次大会「大規模 Web 研究基盤上での自然言語処理・情報検索研究」ワークショップ論文集, pp. 27–20, 2007.

表 10 Wikipedia, ニュース, ブログ中の記述の比較分析 (「ミサイル」)

サブトピック	共通事項	Wikipedia 特有	ニュース特有	ブログ特有
ミサイルの種類	ミサイルの名前とその簡単な説明	多くの種類のミサイルについて、同程度に情報を載せている。ニュースより詳しい。	どのミサイルをどのように使っていくのか、国際情勢から見てそのミサイルを保有するとどう見られるのか。	数を絞ったミサイルの詳細な説明および日本がミサイル防衛をすることの是非を議論
新たな技術	ミサイルに関する新技術について簡単な説明	新技術の紹介、その技術の安全性および研究の現状。	新技術の運用上の問題点	(記事なし)
イージス艦衝突事故	(なし)	(記述なし)	最新式イージス艦の衝突事故において、レーダがミサイルの探知用のものであった。小型船の接近を感知できなかった。	(記事なし)
技術漏洩	(なし)	(記述・記事なし)		MD に関して、日本とアメリカの実験を韓国が見たいといってきたが、日本がそれをきっぱり断った理由は機密漏洩阻止のためだった。
安全保障	他国が保有しているミサイルの性能、および、攻撃された場合の対応策について。	(記述なし)	各国の防衛策の拡張予定	北朝鮮のミサイル発射実験に反応して、日本の MD の重要性を強く主張。「テポドン打ち込まれてまでおとなしくしてる必要は無い」と過激な対応を期待。
MD	日本が導入した防衛ミサイルの性能について。	(記述なし)	日本が MD を行うことによってアジア諸国に与える影響を懸念。	MD に使われるミサイルの仕様や運用方法を事細かに記載。MD がいかに有効なシステムであるかを詳細に解説。
テポドン	(なし)	(記述・記事なし)		北朝鮮を非難。「何らかの方法で制裁を加えたほうがよい」

表 11 Wikipedia, ニュース, ブログ中の記述の比較分析 (「ヒラリー・クリントン」)

サブトピック	共通事項	Wikipedia 特有	ニュース特有	ブログ特有
大統領選以前	(なし)	大統領選出馬前のヒラリーの経歴、および、名前に関するこだわり。	(記事なし)	
スーパーチューズデイ	スーパーチューズデイでの支持率について	(記述なし)	スーパーチューズデイでは接戦を演じたヒラリー氏とオバマ氏だったが、それ以降はヒラリー氏が押されている。	スーパーチューズデイに向けて各陣営が準備を進めている。
女性大統領誕生への期待	(なし)	(記述なし)	一部女性有権者の米初の女性大統領誕生に向けた期待と、メディアの女性差別発言に対する憤り。	(記事なし)
ヒラリー凋落	スーパーチューズデイ以降のヒラリーの支持率凋落について	(記述なし)	スーパーチューズデイ以降のヒラリー陣営の対処の不味さを批判し、ヒラリー氏の苦境を記述。	オバマ氏の政治理念についての考察し、オバマ氏が大統領になると確信。
選挙資金	指名候補争いにおける両陣営の選挙資金の動向について	(記述なし)	公的助成金の受け入れ、および、アメリカ国内の政治資金集めの状況について。	ヒラリー氏、オバマ氏両氏の政治資金の出所についての黒い噂について。
対日本のヒラリーのスタンス	(なし)	(記述・記事なし)		「ヒラリーは大変な親中派であり、日本を軽視している。」
大統領選以降	ヒラリーと國務長官の役職について	オバマ氏からの國務長官就任指名を受け入れた。	(記事なし)	「ヒラリー氏が國務長官に打診されるんじゃないかという憶測が流れている。」

表 12 Wikipedia, ニュース, ブログ中の記述の比較分析 (「バンク・オブ・アメリカ」)

サブトピック	共通事項	Wikipedia 特有	ニュース特有	ブログ特有
合併前のバンクオブアメリカ社の歩み	(なし)	ネーションズバンク社およびバンク・アメリカ社の社歴	(記事なし)	(記事なし)
合併後のバンクオブアメリカ社の歩み	(なし)	バンク・オブ・アメリカ社となってからの社歴	(記事なし)	(記事なし)
株価レポート	市場での株価の動きについてレポート	(記述なし)	主として日本株式市場について、ニューヨーク株式市場の動向についても情報はあがるが、日本の市場との連動について記述.	「市場でバンク・オブ・アメリカの社債保有リスクが上昇. 少なくとも 2001 年 11 月以来の最高に達した。」
米金融不安	サブプライムローン問題や、リーマンブラザーズ倒産に端を発する米金融不安について	(記述なし)	世界を牽引してきた米証券トップ 5 のうち 3 社が市場から淘汰されたり、買収されたりする異常事態について記述.	「資本主義の総本山であるアメリカの銀行も国営化されて金融市場を落ち着かせるしかないのだろう。」「今の日本は外需依存社会だということに、どうも日本の与党の先生方は危機感が足りないというかなんというか……。明日から始まる株式相場の値動きが心配です。」
世界金融市場と日中金融機関	(なし)	(記述なし)	米国発の金融危機に乗じて世界的な金融機関を安値で買収しようとする中国と日本の競争について記述.	(記事なし)
メリルリンチとの合併	バンク・オブ・アメリカ社によるメリルリンチの買収について	メリルリンチを買収した期日や、その際の総額、一株あたりの価格について.	合併に伴い、メリルリンチの株価は上昇したが、バンク・オブ・アメリカの株価は下がった.	「メリルリンチがバンカメによって買収された後、メリルで少なくとも 1 万人が削減される可能性がある。」
米国株式市場	(なし)	(記述なし)	米国のクレジットカード会社 VISA の新規株式公開にともなう利益について記述.	(記事なし)
韓国金融市場	(なし)	(記述なし)	米国のクレジットカード会社 VISA の新規株式公開にともなう韓国経済に対する影響について記述.	(記事なし)