

# ニュース記事の理解支援のための背景知識抽出と補完

田中 祥太郎<sup>1,a)</sup> ヤフト アダム<sup>1,b)</sup> 田中 克己<sup>1,c)</sup>

概要：本稿では，ニュース記事の理解に役立つような背景知識を Web 上のリソースから取得する手法を提案する．ニュース記事の中には，その背景知識がなければ内容を理解しにくいものがある．このようなニュース記事は背景知識の説明を含んでいることがあるが，それらは必ずしも十分であるとは限らない．そこで本稿では，ユーザに対してニュース記事の理解を支援するため，ニュース記事から人物や国家などのエンティティを抽出し，Web 上のリソースからエンティティ間の関係を示す情報を取得することで，記事の内容を補完することを試みる．提案手法では，TextRank アルゴリズムを応用し，記事中のエンティティをランキングする．さらにランキングされたエンティティのうち複数を含むような出来事についての記述を Wikipedia から取得する．実験では，提案手法によるエンティティのランキングと取得された背景知識に対する評価を行う．

キーワード：ニュース，Wikipedia，理解容易性，背景知識

## 1. はじめに

近年のニュースメディアの発達により，日々発行されるニュース記事の数は増加の一途をたどっている．また Web の普及に伴い，特定のトピックに関する記事を主に扱う専門的なニュースメディアも多くなっている．このような記事数やメディア数の増加に伴い，ニュース記事の内容もより詳細で専門的なものとなっている．これらのニュースの中には，その背景知識がなければ理解することが困難なものがある．本研究では，このようなニュースを理解することに役立つ背景知識をユーザに提示することで，ニュースの理解を支援することを目指す．

背景知識が無ければ理解しにくいニュース記事の例を図 1 に示す．この記事は，クリミア地域がウクライナからの離脱とロシアへの併合を表明したことを報じたものである．この記事の内容と深く関わる背景知識として，クリミア地域は 1921 年にはロシア・ソビエト社会主義共和国に所属していたが，1954 年にウクライナ・ソビエト社会主義共和国へ移管され，ソ連の崩壊後もウクライナの支配下に置かれたという事実が挙げられる．

しかし，このような背景知識は必ずしも記事中で説明されているわけではない．また，記事中に説明がある場合でも，複数の時代にまたがる国家や地域間の関係を説明して

### Crimea Votes to Secede From Ukraine as Russian Troops Keep Watch

SIMFEROPOL, Ukraine With thousands of heavily armed Russian troops occupying this perennially embattled peninsula, an overwhelming majority of Crimeans voted on Sunday to secede from Ukraine and join Russia, resolutely carrying out a public referendum that Western leaders had declared illegal and vowed to punish with economic sanctions. ...

(March 16, 2014, The New York Times)

図 1 背景知識を必要とするニュース記事

表 1 図 1 の記事の理解に役立つと考えられる背景知識

年	出来事
1921	ロシアの下にクリミア共和国が設置
1954	クリミア共和国がロシアからウクライナへ移管
1991	ロシアおよびクリミアがソビエト連邦を離脱し独立

いるような文章は，容易に理解できない可能性がある．さらに，このような背景知識が記事中にまとまって記されていない場合，それらを理解することはより難しくなる．

そこで本研究では，ニュース記事の理解に役立つような背景知識を Web 上のリソースから取得する手法を提案する．ニュース記事の理解に役立つ背景知識としては，関連する他のニュース記事や，記事中に登場する人物についての百科事典記事など様々なものが考えられる．本研究では，特にニュース中に現れる人物，国家，地域などのエン

<sup>1</sup> 京都大学大学院情報学研究科  
<sup>a)</sup> stanaka@dl.kuis.kyoto-u.ac.jp  
<sup>b)</sup> adam@dl.kuis.kyoto-u.ac.jp  
<sup>c)</sup> tanaka@dl.kuis.kyoto-u.ac.jp

ティティ間の過去の関係を示す情報に着目する。これは、複数のエンティティ間の時系列的な関係変化は特に理解することが難しいと考えられるためである。図1の記事に対し、本研究で扱うような背景知識の例を表1に示す。

提案手法における入力は、単一のニュース記事とする。提案手法では、まず与えられた記事から背景知識の発見に役立つようなエンティティを抽出する。次に、これらのエンティティをニュース記事での重要度順にランキングする。さらに、ランキングされたエンティティを用いて Web 上のリソースから背景知識を取得し、ニュース記事を補完する。

## 2. 関連研究

ニュース記事などの時系列文書の理解を支援する取り組みとしては以下のような関連研究が挙げられる。Rennison ら [1] は、ニュース記事間の関連性に基づいて各ニュース記事の特徴空間にマッピングし、記事間の関係を可視化する手法を提案した。灘本ら [2] は、あるニュース記事に対し、そのメインピックとサブピックをそれぞれ抽出し、これらの特徴に基づいて関連する過去のニュース記事を取得する手法を提案した。Cui ら [3] は、ニュースのもつ多様な特徴を利用し、ニュース系列における特徴の変化を可視化する手法を提案した。

ニュース記事などの文書からの特徴抽出に関連する取り組みとしては、以下のような関連研究が挙げられる。Mihalcea ら [4] は、文書中の要素間の関係に基づき、各要素をランキングする手法を提案した。この手法は、Brin ら [5] によるリンク解析アルゴリズムである PageRank を応用したものである。Smith ら [6] は、日付や場所などキーワードを用いて非構造化文書からイベントを抽出する手法を提案した。Jatowt ら [7] は、文書中の時間表現やエンティティなどの特徴を用いることで、その文書において注目されている時間軸上の期間を推定する手法を提案した。Milne ら [8] は、文書中の要素から Wikipedia 項目へのリンクを生成する手法を提案した。

ニュース記事などの文書に対する関連情報の発見に関する取り組みとしては、以下のような関連研究が挙げられる。Shahaf ら [9] は、ニュース記事にまたがるトピックに基づき、ニュース記事間の関係を発見し、記事集合を構造化する手法を提案した。高橋ら [10] は、人物や組織などのエンティティの時間的および空間的なインパクトを計算する手法を提案した。

## 3. 背景知識の抽出と補完

背景知識の抽出と補完は次の手順に従って行う。

- (1) エンティティの抽出
- (2) エンティティのランキング
- (3) 背景知識の取得

... Should he annex **Crimea**, **Mr. Putin** could find himself quickly forced into negotiations with the fledgling government in **Kiev** that he has so far refused to recognize or meet, or face a serious conflict over water, energy and other essentials for which **Crimea** is largely dependent on mainland **Ukraine**. ...  
(March 16, 2014, The New York Times)

図2 エンティティ(太字)を含むニュース記事

表2 図2の記事中のエンティティ

語	名称	種別
Crimea	Crimea	地域
Mr. Putin	Vladimir Putin	人物
Kiev	Kiev	地域
Ukraine	Ukraine	国家

### 3.1 エンティティの抽出

エンティティとは、ニュース記事の本文中に現れる人物、組織、出来事、場所、建造物などである。これらは記事が現実世界のどのような事物と関連する話題を扱っているのかを示す重要な特徴である。図2および表2に、ニュース記事とそれに含まれるエンティティの例を示す。本研究では、ニュースを理解するための背景知識として、特にエンティティ間の過去の関係を示す情報に着目している。このため、まずニュース中のエンティティを抽出する必要がある。

エンティティの抽出には2種類の方法が考えられる。1つは、予めエンティティの辞書を用意しておき、ニュース中の各単語をこの辞書と照合する方法である。もう1つは、固有表現抽出などの自然言語処理技術を用いる方法である。前者の利点は、表記の異なる語であっても同一のエンティティとして認識できる点である。欠点は、予め辞書に登録されているエンティティ以外は抽出できない点である。後者の利点は、予め登録されていないエンティティであっても抽出が可能な点である。欠点は、表記が異なる語を別個のエンティティとして認識してしまう点である。

本研究において扱うエンティティは主に実在の国家や人物であるため、これらを登録した辞書を用意することが可能である。また、背景知識を外部リソースから取得するためには、表記が異なる場合でも、指し示す対象が同一であれば同一エンティティとして認識できることが望ましい。これらの理由により、本研究では前者の方法を採用する。実験で使用するツールについては第4章で述べる。

### 3.2 エンティティのランキング

抽出されたエンティティは、背景知識を取得するために利用する。取得する背景知識としては、記事の内容と深く関連する複数のエンティティについて、それらの関係を示

すようなものを想定する。このため、抽出されたエンティティのランキングを行う。ランキングでは、まず記事における重要度が高いエンティティが上位となる必要がある。また、エンティティ間の関係を示す情報を取得するため、重要なエンティティとともに出現するようなエンティティも上位となることが望ましい。

ここで、エンティティを出現回数のみによってランキングした場合、記事の一部のみに頻出するエンティティが上位となる可能性がある。このようなエンティティは、必ずしも記事全体における重要度が高いとは言えない。また、出現回数が少ないエンティティをに対して詳細な順序付けを行うことも困難である。

そこで、エンティティのランキングには、TextRank [4] に類似する手法を用いる。TextRank は、文書の特徴をグラフで表現することで、文書の構成要素をランキングするアルゴリズムである。TextRank によるランキングは、個々の要素の出現頻度だけでなく、要素間の関係を反映したもとなる。すなわち、重要な要素と強く関連する要素が上位にランキングされる。

アルゴリズムの概要は次のようなものである。文書中の単語や文などの要素を節点とし、単語間の共起関係や文の類似度などに基づいて節点間に枝を張ることでグラフを生成する。生成されたグラフは本来無向グラフであるが、各枝を双方向の2つの枝とみなすことで有向グラフを得ることができる。この有向グラフに対して PageRank アルゴリズムを適用することで、各節点をランキングする。

提案手法では、エンティティのランキングに TextRank を応用する。図3に手法のイメージを示す。まず、抽出された各エンティティ  $e_i$  を節点とする。次に、記事のParagraph構造を利用し、同一Paragraphに出現するエンティティ間に枝を張る。すなわち、エンティティ  $e_i, e_j$  間の枝の存在を2値関数  $f(i, j)$  によって表すと、

$$f(i, j) = \begin{cases} 1 & \text{if } \exists p \text{ s.t. } e_i, e_j \in p \\ 0 & \text{otherwise} \end{cases}$$

ここで  $p$  は記事中のあるParagraphとする。

このようにすることで、複数の段落において他の多くのエンティティとともに出現するようなエンティティが上位にランキングされる。また、重要なエンティティとともに出現するエンティティがより上位にランキングされる。

生成された無向グラフを有向グラフに変換し、PageRank アルゴリズムを適用して節点をランキングする手順については通常の TextRank と同様である。

### 3.3 背景知識の取得

背景知識データとしては、過去に発生したイベントの記述を用いる。これらのイベント記述は、ニュースアーカイブや百科事典などから収集する。実験で使用する具体的な

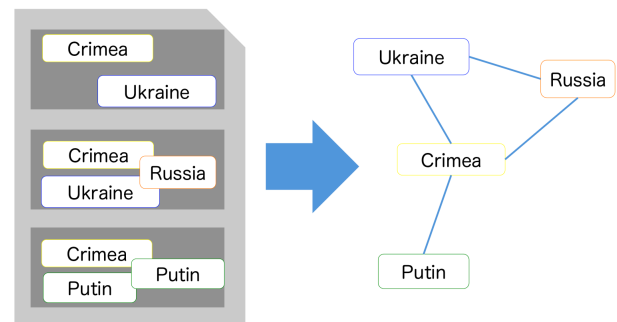


図3 段落構造を応用した TextRank

データについては第4章で述べる。

背景知識の取得は次のように行う。まず収集された各イベント記述に対し、記事からのエンティティ抽出と同様の方法によって予めエンティティを抽出しておく。次に記事から抽出されたエンティティのうち、ランキング上位の  $k$  件を選ぶ。ここで、 $k$  は予め設定した定数である。さらに、収集されたイベント記述のうち、ランキング上位のエンティティを複数含むものを全て取得する。

本稿では主にエンティティのランキングに着目するため、背景知識の取得方法は非常に単純なものとなっている。背景知識の内容に基づくフィルタリングおよびランキング、および取得された背景知識の構造化については今後の課題とする。

## 4. 実験

実験は次の手順に従って行った。

- (1) 背景知識データの準備
- (2) ニュース記事の選定
- (3) エンティティの抽出
- (4) エンティティのランキングの評価
- (5) 取得された背景知識の評価

### 4.1 背景知識データの準備

背景知識の取得元には、オンライン百科事典 Wikipedia 英語版のデータを利用した。Wikipedia には、図4のような西暦の各年に対応する項目ページが存在する。これらの項目ページには、各年に起きたイベントがその日付とともに記されている。実験では、まずこれらのページに記載されているイベント記述を取得し、データベース化した。さらに、各イベント記述から他の Wikipedia 項目へのリンクを抽出し、これらを各イベント記述に出現するエンティティとして扱った。

<p><b>2013</b> - Wikipedia, the free encyclopedia</p> <p>2013 (MMXIII) was a common year starting on Tuesday of the Gregorian calendar, ...</p> <p><b>Events</b></p> <p>January</p> <ul style="list-style-type: none"> <li>January 11 - The French military begins a five-month intervention into the Northern Mali conflict, targeting the militant Islamist Ansar Dine group.</li> <li>January 16-20 - Thirty-nine international workers and one security guard die in a hostage crisis ...</li> </ul>
--

図 4 Wikipedia の各年の項目ページ

表 3 実験に用いたニュース記事

ID	発行日付	タイトル
1	2014-03-16	Crimea Votes to Secede From Ukraine as Russian Troops Keep Watch
2	2014-05-07	Michael Jackson, the Remix
3	2014-05-19	China 's Pollution Challenge
4	2014-05-22	Thailand 's Military Stages Coup, Thwarting Populist Movement
5	2014-06-02	Apple Unveils iOS 8 and OS X Yosemite at Developer Conference
6	2014-06-12	Only a Few Will Decide on What Could Be a Big Shift for Britain
7	2014-06-14	Carrying Fragile Hopes, England Stumbles in Opener
8	2014-06-18	Spain 's Incoming King Takes Over a Throne Heavy With Political Tension
9	2014-06-20	Ruling on Argentina Gives Investors an Upper Hand
10	2014-06-24	Shinzo Abe 's Bid to Shake Up Corporate Japan

#### 4.2 ニュース記事の選定

実験に用いるニュース記事は、The New York Times により 2014 年に発行された記事から 10 件を選んだ。表 3 にその一覧を示す。

選定にあたっては、比較的量の多い記事を中心に選んだ。これは、詳細な内容を含む記事ほど理解する上で背景知識が必要となる可能性が高いという予想に基づくものである。また内容の偏りを防ぐため、政治や社会からエンターテイメントに至るまで幅広いトピックを選んだ。出現するエンティティについてもなるべく偏りのないよう留意した。

#### 4.3 エンティティの抽出

記事からのエンティティ抽出には、Wikipedia Miner<sup>\*1</sup> [8] を利用した。Wikipedia Miner は、与えられた文書から Wikipedia 項目を抽出するツールである。実験では Wikipedia Miner の wikify service を利用し、記事から

<sup>\*1</sup> <http://wikipedia-miner.cms.waikato.ac.nz/>

表 4 エンティティのランキングの評価結果 (nDCG@10)

記事 ID	エンティティ数	出現回数	提案手法
1	37	0.984	1.000
2	19	0.970	0.946
3	18	0.964	0.884
4	15	0.958	0.945
5	17	0.927	0.883
6	13	0.930	0.944
7	14	0.952	0.913
8	23	0.887	0.777
9	19	0.961	0.914
10	30	0.987	0.954
平均	20.5	0.952	0.916

Wikipedia 項目を抽出した。wikify service では、各語が Wikipedia の記事中に出現した場合に対応項目へのリンクが張られる確率を計算できる。実験では、この確率が 0.75 以上と推定された語をエンティティとして抽出した。

wikify service を用いた場合、人物や国家以外に特定の日付や概念なども抽出される可能性がある。本来これらの要素はエンティティとして扱うべきではないが、簡単のため、実験ではこれらも全てエンティティとして扱った。

#### 4.4 エンティティのランキングの評価

エンティティのランキングは第 3 章で述べた手法に従って行った。評価の目的は、提案手法によるランキングが単純な出現回数によるランキングに比べてどの程度有効であるかを確認することである。

ランキングの評価は、3 人の被験者の協力のもとで次のように行った。まず被験者に対し、抽出された各エンティティと記事との関連度を 1 点から 5 点までの 5 段階で評価させた。次に各被験者による各エンティティの評価値の平均を算出し、これを各エンティティの得点とした。さらに、出現回数と提案手法の 2 つの手法によるランキング結果に対し、この得点に基づく nDCG@10 を計算した。

ランキングの評価結果を表 4 に示す。表 4 を見ると、特に量の多いいくつかの記事において、提案手法によるランキングが出現回数によるランキングを上回っていることがわかる。例として、表 5 に ID1 の記事から抽出されたエンティティのランキング結果を示す。これを見ると、この記事において最も重要なエンティティと考えられる Crimea が提案手法では最上位にランキングされていることが確認できる。また、出現回数によるランキングでは差が付けられなかったエンティティについても、提案手法ではより詳細にランキングできていることがわかる。しかしながら、記事全体に対する平均評価値は出現回数のみによる手法を下回った。これは、出現回数のみによるランキングで既に高い評価が得られており、提案手法があまりランキングの向上に寄与できなかったためと考えられる。

表 5 ID1 の記事から抽出されたエンティティのランキング結果

順位	出現回数		提案手法	
	エンティティ	スコア	エンティティ	スコア
1	Russia	26	Crimea	0.113
2	Ukraine	22	Russia	0.095
3	Crimea	21	Ukraine	0.084
4	Crimean Tatars	8	Crimean Tatars	0.037
5	Simferopol	5	Simferopol	0.037
6	Kiev	3	Kiev	0.033
7	Catalonia	2	East Ukraine	0.033
8	Spain	2	Catalonia	0.024
9	East Ukraine	2	Spain	0.024
10	Kosovo	2	Serbia	0.023

表 6 背景知識の評価結果

記事 ID	背景知識数	有用度
1	8	3.00
2	1	4.77
3	20	1.33
4	0	
5	4	1.77
6	4	2.00
7	2	1.33
8	0	
9	0	
10	0	
平均	3.90	2.37

#### 4.5 取得された背景知識の評価

背景知識の取得は第 3 章で述べた手法に従って行った。評価の目的は、ランキング上位のエンティティのみを用いた場合、取得される背景知識が記事の理解にどの程度有用であるかを確認することである。

背景知識の取得には、各記事から抽出されたエンティティのうち上位の 10 件を用いた。評価については、エンティティのランキングと同様に 3 人の被験者の協力のもとで行った。まず被験者に対し、取得された各背景知識が記事を理解する際にどの程度役立つかを 1 点から 5 点の 5 段階で評価させた。次に各被験者による各背景知識の評価値の平均を算出し、これを各背景知識の得点とした。さらに、各記事に対して取得された背景知識の得点の平均を算出し、これを各記事に対する背景知識集合の有用度とした。

取得された背景知識の評価結果を表 6 に示す。これを見ると、いくつかの記事においては、取得された背景知識集合の有用度は高くなっていることがわかる。しかしながら、平均の有用度は 2.30 と、あまり高くなかった。また、取得された背景知識の数は平均で 3.90 と少なく、取得された背景知識がひとつもなかった記事も存在した。原因としては、背景知識として利用したイベントの記述が簡

表 7 ID1 の記事に対して取得された背景知識

日付	出来事
1992-02-14	Ukraine and 4 other nations in the Commonwealth of Independent States reject Russia's proposal to maintain unified armed forces. Ukraine, Moldova, and Azerbaijan announce they will go ahead with plans to create their own armed forces.
1992-05-05	Russian leaders in Crimea declare their separation from Ukraine as a new republic. They withdraw the secession on May 10.
2007-01-08	Russian oil supplies to Poland, Germany, and Ukraine are cut as the Russia-Belarus energy dispute escalates; they are restored 3 days later.
2009-01-07	Russia shuts off all gas supplies to Europe through Ukraine. Prime Minister Vladimir Putin publicly endorses the move and urges greater international involvement in the energy dispute.
2013-11-21	Euromaidan pro-EU demonstrations begin in Ukraine after President Viktor Yanukovich rejects an economic association agreement between the European Union and Ukraine in favor of closer ties to Russia.

表 8 ID1 の記事に対して取得されなかった背景知識

日付	出来事
1954-02-19	1954 transfer of Crimea: The Soviet Union transfers the Crimean Oblast from the Russian SFSR to the Ukrainian SSR.

潔なものであり、そこから抽出できたエンティティが少数であったことが考えられる。また、今回は記事から抽出したエンティティのみを用いて背景知識を取得したが、記事に現れない重要エンティティの発見手法についても今後考慮すべきと考えられる。例として、表 7 に ID1 の記事に対して取得された背景知識を示す。また、表 8 にデータベース中には存在したが取得されなかった背景知識を示す。表 8 に挙げた背景知識を取得できなかったのは、Russian SSR や Ukrainian SSR といったエンティティが、記事中の Russia や Ukraine といったエンティティとは別の項目として Wikipedia に存在するため、別のエンティティとして認識されたことが原因と考えられる。

#### 4.6 考察

一連の実験の結果、提案手法によるエンティティのランキングはいくつかの場合において有効であるが、提案手法によりランキングされたエンティティのみでは有用な背景知識を取得できないことが明らかになった。

今後の課題としては、主に次のようなものが挙げられる。

1つ目は、エンティティの抽出手法の改善である。今回エンティティの抽出には Wikipedia Miner を用いた。しかしこの方法では、Wikipedia に存在する項目が全て抽出対象となる。このため、提案手法でエンティティとして扱う人物、国家、地域など以外の要素が抽出されることがある。この点については、何らかの方法でエンティティ以外の要素を除外することが必要と考えられる。

2つ目は、エンティティのランキング手法の改善である。実験では、提案手法によるランキングが、出現回数のみによるランキングに比べ、いくつかの場合で優位であることが確認できた。しかし、提案手法によるランキングは依然としてエンティティの出現回数に依存しているため、出現回数が少ないが重要であるエンティティを発見することは難しい。この点については、エンティティそのものではなく、エンティティ間の関連をランキングすることが考えられる。また、提案手法では記事の段落構造を利用するため、段落数が少ない記事に対しては精度が著しく低下する。この点についても改善が必要である。

3つ目は、背景知識の取得手法の改善である。提案手法では、上位にランキングされたエンティティのみを背景知識の取得に用いた。しかし、上位にランキングされなかったエンティティや、記事中に全く出現しないようなエンティティでも、背景知識の取得に有用であるものが存在する可能性がある。これらについては、取得された背景知識からさらにエンティティを抽出し、これを用いて再帰的に背景知識を取得することなどが考えられる。また、今回は背景知識を取得するための特徴としてエンティティのみを用いたが、ニュース記事から得られるエンティティ以外の特徴を利用することも考えられる。

## 5. おわりに

本稿では、ニュース記事の理解に役立つような背景知識を Web 上のリソースから取得するための手法を提案した。提案手法では、まずニュース記事から背景知識の取得に利用するエンティティを抽出した。次にこれらをニュース全体における重要度に基づいてランキングした。ランキング手法としては、TextRank に記事中の段落構造を応用したものを利用した。さらに、上位にランキングされたエンティティのうち複数を含むようなイベント記述を Wikipedia の各年の項目ページから取得した。実験では、実際のニュース記事を用いてエンティティのランキング精度と取得された背景知識の適合度を評価した。

今後の課題としては、エンティティの抽出手法の改善、エンティティのランキング手法の改善、背景知識の取得方法の改善などが挙げられる。また、実験における記事数、被験者、背景知識データの量が十分とは言えないため、今後はより大規模な実験により評価を行うことも考えたい。

将来の展望としては、取得された背景知識をユーザに提

示する際、どのような方法が有効であるかを検討する必要がある。本稿では、主にエンティティの抽出と背景知識の取得に取り組んだ。しかし、文章で記述された背景知識をそのままユーザに提示することは、ニュースの理解を支援する上で必ずしも最良の方法とはいえない。今後は取得された背景知識を構造化し、ユーザが理解しやすい形で視覚化して提示するといったことにも取り組みたい。

## 謝辞

本研究の一部は、文科省科研費基盤 (A) 「ウェブ検索の意図検出と多角的検索意図指標にもとづく検索方式の研究」(24240013, 研究代表者: 田中克己)、戦略的創造研究推進事業 (さきがけ) 「集合記憶の分析および歴史文書からの知識抽出手法の開発」(研究代表者: Adam Jatowt) によるものです。ここに記して謝意を表します。

## 参考文献

- [1] Rennison, E.: Galaxy of news: An approach to visualizing and understanding expansive news landscapes, *Proceedings of the 7th annual ACM symposium on User interface software and technology*, ACM, pp. 3–12 (1994).
- [2] Nadamoto, A. and Tanaka, K.: Time-based contextualized-news browser (t-cnb), *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, ACM, pp. 458–459 (2004).
- [3] Cui, W., Qu, H., Zhou, H., Zhang, W. and Skiena, S.: Watch the story unfold with textwheel: Visualization of large-scale news streams, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 3, No. 2, p. 20 (2012).
- [4] Mihalcea, R. and Tarau, P.: TextRank: Bringing order into texts, *Proceedings of the 2004 Conference of the Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 404–411 (2004).
- [5] Brin, S. and Page, L.: The anatomy of a large-scale hypertextual Web search engine, *Computer networks and ISDN systems*, Vol. 30, No. 1, pp. 107–117 (1998).
- [6] Smith, D. A.: Detecting and browsing events in unstructured text, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 73–80 (2002).
- [7] Jatowt, A., Au Yeung, C.-M. and Tanaka, K.: Estimating document focus time, *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, ACM, pp. 2273–2278 (2013).
- [8] Milne, D. and Witten, I. H.: An open-source toolkit for mining Wikipedia, *Artificial Intelligence*, Vol. 194, pp. 222–239 (2013).
- [9] Shahaf, D. and Guestrin, C.: Connecting the dots between news articles, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 623–632 (2010).
- [10] 高橋侑久, 大島裕明, 山本光穂, 岩崎弘利, 小山 聡, 田中克己: インパクトを考慮した歴史エンティティの重要度計算手法, *情報処理学会論文誌*, Vol. 52, No. 12, pp. 3542–3557 (2011).