

# 段階的なテーマ詳細化によるニュース情報の体系的獲得

## Acquiring Organized Information from News by Incremental Theme Refinements

谷口 祐太郎<sup>1\*</sup> 小林 哲則<sup>1</sup> 林 良彦<sup>1</sup>  
Yutaro Taniguchi<sup>1</sup> Tetsunori Kobayashi<sup>1</sup> Yoshihiko Hayashi<sup>1</sup>

<sup>1</sup> 早稲田大学 理工学術院

<sup>1</sup> School of Science and Engineering, Waseda University

### Abstract:

We propose an interactive system which allows a user to efficiently acquire organized information from a news corpus. Initiated by a topic specified by a user, the system first extracts a relevant set of sentences, from the corpus and divides them into themes each designated by an additional query term. The user then can select one of them, and the system maintains an interactive information access session by incrementally refining the search queries. This incremental process enables user's topic-specific yet efficient information access activities. Small-scaled experiments confirmed that the proposed system could provide the user with more effective supports than an ordinary search engine in the sense that it concisely provides a set of chronologically-sorted sentences classified by the themes. The experiments however further identified important issues, including the reduction of duplicate themes and sentences, and the deletion of sentences that cannot tell their meaning without neighboring contexts. For future work, we first implement automatic title generation, and then complete a total system.

## 1 はじめに

### 1.1 研究背景と目的

ニュースから体系的かつ効率的に情報を取得するために、ニュースコーパスからトピック（知りたい事柄）に関連する複数のテーマを抽出し、各テーマに関連する文が時系列順に並んだ文集合（ストーリーライン）を出力するシステムを提案する。

情報が個人や団体の資源になるといわれるなかで、世間にあふれている情報を得る労力を削減することは非常に重要である。とくに、ニュースの情報を得ることで世間の事件や事象を把握することは、日常会話の話題提示や、経済の未来予測、自らの意見の立ち位置の明確化に非常に有用である。ある事柄についてニュースから情報を取得しようとした場合、事柄に関するニュース記事群をそのまま読むと、事柄についての様々なテーマが混在しており量も多いため、非常に労力を要する。一方で、「Wikipedia」のような、ある事柄に関しての情報のまとめは、ニュースをもとに加工された、編集

者の主観が含まれているような情報であり、客観性や公平性を望むことができないという問題がある。

したがって、ニュースから体系的かつ効率的に無加工の情報を獲得することは価値があると考えられる。

### 1.2 本研究の位置づけ

本稿では、ニュース記事集合とユーザが知りたい事柄を入力とし、テーマごとのタイトル付きの複数の文集合を出力とするシステムを提案するが、ニュースからの体系的・効率的な情報獲得の既存手法はさまざまである。

例えば、「ニュース検索結果を記事単位でクラスタリングした結果を見る」[2][3]、「ニュース全体をもとにテーマを抽出し、テーマごとの記事集合を図示する」[1]といった試みがある。また、「文書に対して、ユーザが指定したテーマに基づいて文を抽出する」[5]という試みを、ニュース記事全文をひとつの文書とみなしニュース記事に適用すれば、ニュースからの体系的・効率的な情報獲得につながる。これらの研究と本研究での提案手法を入出力の形式に着目してまとめたものを表1に示す。「タイトルを表す語」とは、「出力された各文集

\*連絡先：早稲田大学 理工学術院  
〒162-0042 東京都新宿区早稲田町 27 早稲田大学 40  
号館 701 号室  
E-mail: taniguchi@pcl.cs.waseda.ac.jp

合または文書集合に、その内容を表すタイトルがつけられているかどうか」を示している。ここで、本研究と先行研究との差異に着目し、以下の4点に着目することで、提案システムを位置づける。

1. 入力にはユーザの意向を表す語が含まれる
2. 出力される集合の要素は文書ではなく文である
3. 出力される文集合または文書集合は複数ある
4. 各文集合または文書集合にタイトルを表す語がついている

**入力にはユーザの意向を表す語が含まれる** ニュースから情報を取得するさい、ユーザが指定するトピックを入力として受け付けることで、ユーザの知りたい事柄を出力に反映させる。

**出力される集合の要素は文書ではなく文である** ニュースからの情報取得の効率を考えると、適切な文集合が出力されるならば、文書集合より文集合を読むほうが好ましいと考えられる。あるトピックについての情報を取得するさい、同じ記事の中でもトピックに関連する文と関連しない文が存在し、必ずしも記事全文を読む必要はないため、文単位での集合を見るほうがよいとする。

**出力される文集合または文書集合は複数ある** ニュース記事群の中には、トピックに関するさまざまなテーマが混在している。例えば、「オバマ」というトピックの「オバマの広島訪問」、「オバマの政策」、「オバマ政権への評価」といったテーマを別々に見ることで、ニュースを体系的に読むことが可能となる。

**各文集合または文書集合にタイトルを表す語がついている** 出力された複数の文集合または文書集合をユーザが実際に読むときには、ユーザは自分が読みたいと思った集合を選択するが、このときタイトルがつけられていれば、選択の支援となる。

## 2 提案システムとインターフェイス

### 2.1 システムの提案

ニュースから体系的かつ効率的に情報を取得するために、1.2節で述べたとおり、「入力にはユーザの意向を表す語が含まれる」、「出力される集合の要素は文である」、「出力される文集合は複数ある」、「各文集合にタイトルを表す語がついている」という要素を満たすように、知りたい事柄を与えるとニュースコーパスから

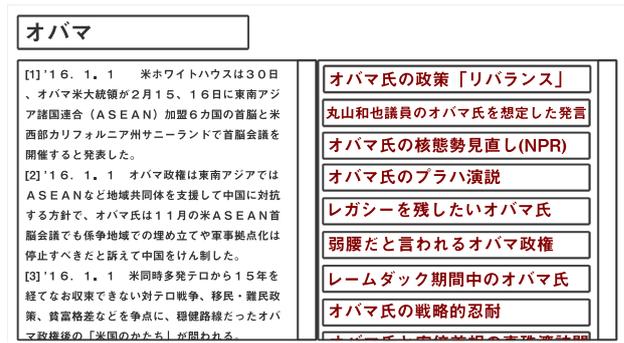


図 1: 想定インターフェイス (階層 1)

この事柄に関連するテーマを抽出し、テーマごとに関連する文集合を文単位で抽出し、時系列に沿って提示するシステムを提案する。

### 2.2 想定インターフェイスと必要事項

#### 2.2.1 想定インターフェイス

提案システムを達成するためのインターフェイスを図1、図2に示す。ここでは、知りたい事柄(トピック)として「オバマ」が与えられた場合を例としている。以下に、想定する使用手順を示す。

1. 初期クエリとして、トピックである「オバマ」を与える。(図1)
2. 左部分に、ニュースコーパスから検索された、「オバマ」に関連する文が時系列順に並んだ「ストーリーライン」が表示される。右部分に、「オバマ」に関連する文集合から取得された「テーマ」が表示され、各テーマにはタイトルがつけられている。(図1) 本稿では、このタイトルを「テーマタイトル」を呼ぶ。
3. 右部分のテーマタイトルのうち一つ、例えば「オバマ氏のプラハ演説」を選択すると、左部分には、前段階から抽出された「オバマ」に関連する文集合からさらに抽出された、「オバマ氏のプラハ演説」に関連する文が、右部分にはその文集合からさらに取得されたテーマが表示される。(図2)
4. ユーザは、興味のあるテーマを選択し、そのテーマに関連する文集合を読むことで、初期クエリに関する様々なテーマに基づいた情報を取得することができる。

表 1: 提案システムと先行研究の比較

	入力	出力		
	形式	形式	複数であるか	タイトルを表す語
提案システム	文書集合 + 語	文集合	○	○
[2]	文書集合	文書集合	○	×
[1]	文書集合	文書集合	○	○
[5]	文集合 + 語	文集合	×	×

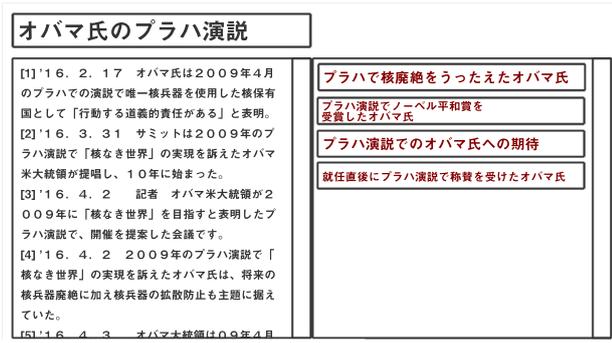


図 2: 想定インターフェイス (階層 2)

## 2.2.2 想定インターフェイスの必要事項

2.2.1 節で提示したインターフェイスを達成するための必要事項を 3 点, 以下に示す.

- 適切なテーマ取得  
 文集合を与えると, ある一つのテーマに沿った複数の文集合が導かれるような処理.
- テーマタイトル生成  
 ユーザがテーマを選択するきっかけとなるテーマタイトルの生成.
- 前階層からの適切な文集合抽出  
 各文集合についての, その文集合のテーマに関連する文の, 前階層からの適切な抽出.

## 3 テーマ取得・文集合抽出

2.2.2 節で示した必要事項のうち, 「(1) 適切なテーマ取得」と「(3) 前階層からの適切な文集合抽出」を達成するための手続きを検討し実装した.

### 3.1 テーマ取得・文集合抽出の提案手法

2.2.2 節で示した必要事項 (1) テーマ取得, 必要事項 (3) 文集合抽出を実現するための手法を提案する. 提案

### Algorithm 1 テーマに沿った複数の文集合の抽出方法

**Input:** ニュース文集合  $S$ , 初期クエリ  $q$ , 階層数  $d$ , 重要語数  $N$ , 閾値  $t$  (最少文数)

- $First \leftarrow \{s | s \text{ は } S \text{ のうち } q \text{ を含む文}\}$
- $OutSets \leftarrow \phi$
- $NewSets \leftarrow \{(q, First)\}$
- $OutSets \leftarrow OutSets \cup NewSet$
- for** from 0 to  $d-1$  **do**
- $PreviousSets \leftarrow NewSets$
- $NewSets \leftarrow \phi$
- for all**  $set$  in  $PreviousSets$  **do**
- $Temp \leftarrow Algorithm2(set[0], set[1], N, t)$
- $NewSets \leftarrow NewSets \cup Temp$
- end for**
- $OutSets \leftarrow OutSets \cup NewSets$
- end for**

**Output:**  $OutSets$

手法を図示したものを図 3 に示す. 重要語を取得することで, 適切なテーマ取得を行い, さらにその重要語を検索式に追加することで, 適切な文集合を抽出するという狙いである. 文の集合をもとに, 重要語をつなげた検索式をユーザに提示する研究としては, 松生ら [7] の研究が挙げられる. ただし, これは Web ページを対象として, ユーザが自身の検索する目的を明らかにするためにキーワード式を提示するという動機である. それに対して本研究は, ユーザが知りたい事柄に対して, 様々な側面から情報を取得するという狙いを狙っている.

本手法は, ユーザのテーマ選択に応じて文集合が生成されるのではなく, あらかじめ一定階層  $d$  までの文集合を生成しておいて, ユーザの選択に応じてその文集合を表示するという想定である. 擬似コード 1 で示したアルゴリズムによって, 検索式とこの検索式により得られる文集合のタプルの集合を生成する. なお, 擬似コード 2 のアルゴリズムを, 擬似コード 1 で参照している.

知りたい事柄としてユーザが入力する初期クエリを  $q$  とする.

**Algorithm 2** 検索式と文集合のセット (Q, S) からの  
 テーマに沿った複数の文集合の抽出方法

**Input:** 検索式 Q, 文集合 S, 重要語数 N, 閾値 t(最  
 少文数)

- 1:  $OutSets \leftarrow \phi$
- 2:  $K \leftarrow \{k | k \text{ は } S \text{ に含まれる単語のうち 1式で示した  
 手法で重要語抽出したときの上位 } N \text{ 語}\}$
- 3: **for all**  $k$  in  $K$  **do**
- 4:  $Temp \leftarrow \{s | s \text{ は } S \text{ のうち } k \text{ を含む文}\}$
- 5: **if**  $|Temp| \geq t$  **then**
- 6:  $NewQ \leftarrow Q \cup k$
- 7:  $TempSet \leftarrow (NewQ, Temp)$
- 8:  $OutSets \leftarrow OutSets \cup TempSet$
- 9: **end if**
- 10: **end for**

**Output:**  $OutSets$

1. ニュース文集合から, q を含む文を抽出し, 文集合  $S(q)$  とする.
2. q を含む文集合に含まれる単語を, 重要度が大きい順に N 語抽出する. この重要語を  $(x_1, x_2, \dots, x_N)$  とする. なお, 重要度の指標は, 式 (1) を用いた.

$$R_{morph} = \frac{c_{morph}}{c_{all}} \cdot \frac{C_{all}}{C_{morph}} \quad (1)$$

ただし, 式 (1) の  $R_{morph}$  は, 単語 morph の重要度,  $c_{morph}$  は文集合における morph の出現回数,  $c_{all}$  は文集合における全単語の出現回数,  $C_{morph}$  はコーパス全体における morph の出現回数,  $C_{all}$  はコーパス全体における全単語の出現回数を示す. 文集合における morph の出現頻度を, コーパス全体における morph の出現頻度で正規化したといえる. コーパス全体における morph の出現頻度の log をとった式で重要度を測る場合もある [4] が, 本研究に適用したところ, 出現頻度が少ない単語が上位に来たため, log を使わない式を用いることとした.

重要語を抽出するさい, 自立語のみという制約を設けた. また, 単語をカウントするさい, 上の階層で既に取得した検索式に含まれる任意のキーワード (初期クエリ含む) との距離 (原文における単語間の距離) が 3 以下, 上の階層で既に取得した検索式に含まれる任意のキーワード (初期クエリ含む) との間に助詞が含まれる, という場合のみカウントするようにした. これは, 前段階のキーワードと同格にあるような語を重要語に含めないという狙いがある. 例えば, 「オバマ」に対して「大統領」というキーワードは, 多く共起

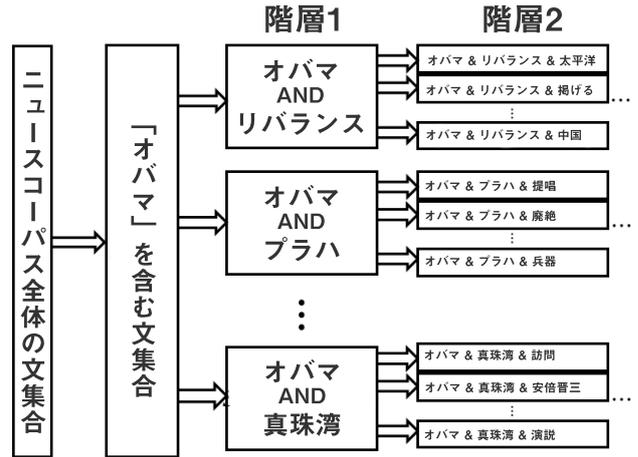


図 3: 文集合抽出の提案手法

するが, 文集合を限定する意味が少ない. また, 「する」「いる」「こと」「ない」は, 自立語であるが意味の少ない語とみなし, stop words とした.

3. 「q AND  $x_n$ 」を含む文を抽出し, それぞれの文集合を  $S(q \text{ AND } x_n)$  とする. これを  $1 \leq n \leq N$  について行う. (階層 1 の文集合)
4. 「q AND  $x_n$ 」の文集合に含まれる単語から, さらに重要語を抽出する. このとき, 重要度が大きい順に N 語抽出する. この重要語を  $(x_{n1}, \dots, x_{nN})$  とする. これを  $1 \leq n \leq N$  について行う.
5. 「q AND  $x_n$  AND  $x_{nm}$ 」を含む文を抽出し, それぞれの文集合を  $S(q \text{ AND } x_n \text{ AND } x_{nm})$  とする. これを  $1 \leq m \leq N, 1 \leq n \leq N$  について行う. (階層 2 の文集合)
6. 以上を階層 d まで繰り返し, 各検索式と検索式から導かれる文集合のタプルの集合を出力とする. ただし, 文集合に含まれる文の数が t 未満のときは, 出力に含めない.

## 3.2 テーマ取得・文集合抽出の実装

### 3.2.1 前処理

2016 年の毎日新聞の記事が収録されたコーパス「CD-毎日新聞 2016 データ集」の各記事データの「日付」と「本文」を利用して実験を行った. 3.1 節で示した提案手法を実現するために, 以下の項目が満たされるように前処理を行った.

1. 各記事データの本文が, 文単位に分割されている.

2. 分割された各文が、単語単位に分割されている。  
また、各単語が自立語であるかどうかの情報を保持している。

**記事本文の文分割** 各記事に対して、記事本文を入力として、文単位に分割されたリストを出力とする。文分割は、以下のようなルールで行った。

1. ”。”, ”!”, ”?”のいずれかと、その次の文字の間で分割する。
2. ただし、カギ括弧内では分割しない。
3. パラグラフの末尾は必ず分割する。

**各文の単語分割** 各文に対して、単語単位に分割されたリストを出力とする。単語分割には、形態素解析システム Juman および構文解析システム KNP の固有表現を特定する機能を用いた。Juman の形態素解析による一形態素を一単語をみなす。ただし、人、場所の固有表現と解析された場合は、形態素としては分割されていても、一つの固有表現を一単語とみなす。(例:「安倍」「晋三」→「安倍晋三」) また、Juman の解析による自立語であるかどうかの情報を保持する。

### 3.2.2 提案手法の実装

3.1 節で示した提案手法の実装について述べる。

初期クエリ  $q$ 、最大階層  $d$ 、取得重要語数  $N$ 、文数閾値  $t$  を入力として、階層数、検索式、文集合を値として持つ dict のリストを出力する。ここで、同階層で文集合が完全に一致するものはマージし、各検索式の最後のキーワードを両方保持する。例えば、「アベノミクス、税収、1億」を検索式にもつ文集合と「アベノミクス、税収、活躍」を検索式にもつ文集合が完全に一致した場合、「アベノミクス、税収、(1億、活躍)」を検索式にもつ文集合として出力する。また、類似度が95%以上である文のペアについて時系列が早いほうの文のみを結果に含めることで、文の重複を回避する。

## 4 提案システムの評価実験

提案システムとインターフェイスの評価実験を行う。[6] では、インターフェイスの必要事項のうち「文集合抽出」に着目し、文集合の正解データを作成することで、取得された各テーマに対してどれだけ正確に文が抽出されたかについての評価を行ったのに対し、本稿ではインターフェイスのモックアップを作成し被験者に利用させることで、提案システムとインターフェイス全体についての評価を行う。

## 4.1 実験手法

### 4.1.1 提案システムの準備

2.2.2 節で示したインターフェイスの必要事項のうち、「テーマ取得」と「文集合抽出」を 3.1 節で示した手法で実装し、「タイトル生成」を手動で理想的なタイトルをつけることで実現する。あらかじめ定めたトピック(「オバマ」、「EU」、「東芝」)それぞれを入力として、「CD-毎日新聞 2016 データ集」をコーパスとしたときの出力結果から、インターフェイスのモックアップを自動生成する(提案システム)。このときのシステムの実行条件を、表 2 に示す。図 4 に、「オバマ」をトピックとするモックアップの画面を示す。

### 4.1.2 ベースラインシステムの準備

オープンソースの全文検索サーバー「Fess」<sup>1</sup>を用いて、「CD-毎日新聞 2016 データ集」を自由なクエリで記事単位検索できるようにする(ベースラインシステム)。検索エンジンはニュースなどから情報を獲得する目的で日常的に利用されており、提案システムの比較対象として自然であると考えた。

### 4.1.3 実験内容

提案システムとベースラインシステムをそれぞれ用いて、3つのトピックに関する情報を調べて、トピックごとにまとめを作成することを5人の被験者に要求する。ここでいうまとめの作成とは、利用者が獲得した情報の中で有益だと思うものについて記述するという意味であり、高度な要約などは要求していない。まとめを作成するさいに、ニュース文からコピーアンドペーストすることを許す。

評価実験の公平性を保つために、被験者 A~E のうち、被験者 A, C, E は、「オバマ」、「東芝」について調べるときは提案システムを先に利用し、「EU」について調べるときはベースラインシステムを先に利用するようにし、被験者 B, D は逆の順番に2つのシステムを利用するように実験を設計した。

実験中に、被験者は実験者に質問・意見・感想を自由に述べる。最後に被験者は、「ニュースから体系的に情報を獲得できたのはどちらのシステムか」、「ニュースから効率的に情報を獲得できたのはどちらのシステムか」という2つの質問に対して、「提案システム」、「どちらか」と提案システム、「どちらともいえない」、「どちらか」とベースラインシステム、「ベースラインシステム」の5つからそれぞれ解答を選択する。

<sup>1</sup><https://github.com/codelibs/fess/releases/tag/fess-11.4.9>

表 2: モックアップ作成時のシステム実行条件

使用コーパス	「CD-毎日新聞 2016 データ集」
トピック	「オバマ」, 「EU」, 「東芝」
取得重要語数	300
階層数 d	3
文数閾値 t	5

表 3: 体系的・効率的であるかの質問に対する被験者の回答結果

被験者	回答	
	質問 a	質問 b
A	1	5
B	4	5
C	5	5
D	4	2
E	5	3

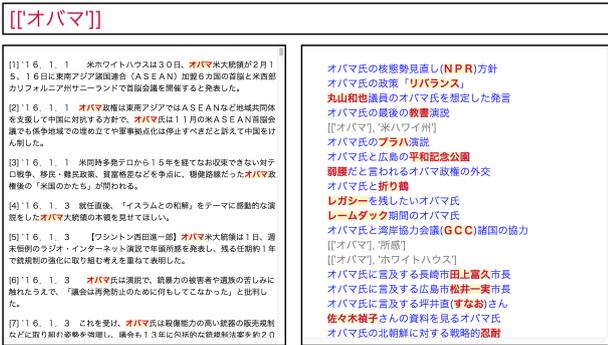


図 4: システムのモックアップ

システム」の5つからそれぞれ解答を選択する。回答結果を表3に示す。これによると、5人中4人が質問aに対して4または5と回答し、5人中3人が質問bに対して4または5と回答しているため、提案システムはニュースから体系的かつ効率的に情報を取得することに貢献しているといえる。なお、被験者Aが質問aに対して1と回答したのは、「提案システムは、提示されたテーマがトピックに関する情報を網羅している保証がないため、体系的とはいえない」との理由である。

## 4.2 実験結果と議論

### 4.2.1 トピックに関して作成されたまとめについて

図5に、被験者Aによる「オバマ」に関するまとめ(ベースラインシステム利用)、図6に、被験者Aによる「オバマ」に関するまとめ(提案システム利用)を示す。各被験者のまとめを比較すると、システムの特性上、ベースラインシステムよりも提案システムのほうが、まとめる観点が被験者によって一定している傾向にある。また、例えばベースラインシステムを利用した「オバマ」に関するまとめでは、5人中3人の被験者が、オバマ氏の後任であるドナルド・トランプ氏について言及しているが、提案システムが提示するテーマにはトランプ氏に関するものは含まれていない。すなわち、提案システムには、多くの利用者にとって重要なテーマであっても表示されないものがあるという問題点があり、テーマ取得の方法(重要語の指標)に改善が要求されていると考える。

### 4.2.2 質問回答結果について

各被験者は実験終了後に、「(質問 a) より体系的にニュースから情報を獲得できたのはどちらのシステムか。」「(質問 b) より効率的にニュースから情報を獲得できたのはどちらのシステムか。」という2つの質問に対して、「1. ベースラインシステム」、「2. どちらかというベースラインシステム」、「3. どちらともいえない」、「4. どちらかという提案システム」、「5. 提案

### 4.2.3 被験者の意見・感想

各被験者は、実験の最中に、自由に各システムに対する意見・感想を述べる事ができる。以下に、意見・感想から導くことができる提案システムの問題点を示す。

**提示されるテーマの問題** 1人の被験者が、各トピックに関して提示されたテーマがトピックのすべての情報を網羅しているとは限らないということを指摘した。1人の被験者が、提示されたテーマに偏りがあるのではないかと述べた。1人の被験者が、提示されたテーマがトピックに関して重要なテーマであることの保証がないということを指摘した。

**文集合内の文の冗長性** 2人の被験者が、同じ文集合内に似たような文が含まれていて冗長に感じるという意見を述べた。3.2.2節で述べた通り、重複度が95%以上の文のペアは、一方を削除することで重複を解消しているが、重複度の閾値の決め方に改善が要求されていることを示す。

**文集合間の文の冗長性** 2人の被験者が、異なるテーマの文集合であるにもかかわらず、含まれる文がほとんど同じであるため冗長に感じるという意見を述べた。3.2.2節で述べた通り、含まれる文集合が完全に一致する文集合はマージするようにしているが、例えばある2つの文集合のJaccard係数が一定以上の場合にはマー

ジするといったように、マージする条件の緩和が要求されていることを示す。

**テーマの階層数の設定が不十分である** 2人の被験者が、さらに階層を進んで文を絞りたいたいと思ったときに、表示されるテーマがないという問題を指摘した。階層数や文数閾値の決め方に改善が要求されていることを示す。

**文単体では意味が通らないことがある** 3人の被験者が、文集合に含まれる文の中には、主語が省略されているなどの理由から、文単体では意味がわからないものがあるという指摘をした。この問題の解決策の一つとして、インターフェイス上で文を選択すると、文を抽出する元となった記事を表示させるようにするというものが挙げられる。また、抽出された文の前後の文を、意味が通る最少の文数だけ表示させることで、効率性を維持することができると思う。

以上の問題点の他に、2人の被験者が、1.2節で述べたシステムの特徴である「文単位で出力されること」に対して、有用であると言及した。3人の被験者が、「テーマごとに出力されること」に対して、有用であると言及した。また、1人の被験者が、2.2.2節で述べたインターフェイスの必要事項である「タイトルがつけられていること」に対して、有用であると言及した。2人の被験者が、文が時系列順に並んでいることに対して、有用であると言及した。

## 5 まとめ・今後の計画

知りたい事柄(トピック)を与えるとニュースコーパスからこのトピックに関連するテーマを抽出し、関連する文集合を提示するシステムを提案し、これを達成するための必要事項(「テーマ取得」「タイトル生成」「文集合抽出」)を提示した。そのうち「テーマ取得」「文集合抽出」に着目し、段階的に重要語を検索式に追加する手法によって達成を試みた。提案システムの評価手法を提案し、評価を行うことで、提案システムが目的を達成していることを示した。また、被験者からの意見・感想を求めることで、実際に利用者が感じる提案システムの有用性と問題点が明らかとなった。

今後は、本稿では手動で実現した「タイトル生成」を自動で行う手法を提案し、インターフェイスを全自動化する。実際に利用者が自由に検索できるように提案システムを実装する。また、実験により明らかになった問題点の改善手法を検討する。一方で、明確な評価基準が存在しないようなシステムに対してどのような評価手段がありうるかについての議論を深める。

## 参考文献

- [1] Philippe Laban, Marti A. Hearst, "newsLens: building and visualizing long-ranging news stories", Proceedings of the Events and Stories in the News Workshop, pp. 1-9, 2017.
- [2] Srinivas Vadrevu, Choon Hui Teo, Suju Rajan, Kunal Punera, Byron Dom, Alex Smola, Yi Chang, Zhaohui Zheng, "Scalable Clustering of News Search Results", Proceedings of the fourth ACM international conference on Web search and data mining, pp. 675-684, 2011.
- [3] Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma, Jinwen Ma, "Learning to Cluster Web Search Results", Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 210-217, 2004.
- [4] 大竹清敬, 岡本大吾, 児玉充, 増山繁, "重要文抽出, 自由作成要約に対応した新聞記事要約システム YELLOW", 情報処理学会論文誌, Vol.43, pp. 37-47, 2002.
- [5] 砂山渡, 谷内田正彦, "観点に基づいて重要文を抽出する展望台システムとそのサーチエンジンへの実装", 人工知能学会論文誌, Vol.17, pp. 14-22, 2002.
- [6] 谷口祐太郎, 小林哲則, 林良彦, "ニュースからのトピックに関するストーリーラインの生成", 言語処理学会, 第24回年次大会, 2018.
- [7] 松生泰典, 是津耕司, 小山聡, 田中克己, "検索結果の概要を表すキーワード式生成による質問修正支援", 電子情報通信学会データ工学ワークショップ (DEW2005), 1C-i9, 2005.

- 自らの選挙戦で「チェンジ (変革)」、そして「イエス・ウィ・キャン (やればできる)」のスローガンを掲げて華々しく登場。2009年1月、黒人初の米大統領に就任
- 就任間もなく「核兵器なき世界」の実現を訴えノーベル平和賞を受賞したが、外交、安保、内政とも理想と現実のはざままで揺れた2期8年だった。
- イラクとアフガニスタンでの二つの戦争の「終結」を公約に掲げ、2011年末にイラクからの軍撤退を果たしたものの、過激派組織「イスラム国」(IS)の台頭により、14年には再派遣に追い込まれ、戦争終結は実現できなかった
- 5/27, 原爆投下国のトップとして初めて被爆地・広島を訪問し、平和記念公園(広島市中区)での演説で改めて「核兵器なき世界」を訴えた
- 12/9の会見で、大統領選への干渉を狙ったサイバー攻撃について再調査するよう情報機関に指示
- イラク戦争終結や医療保険制度改革(オバマケア)など多くの公約を達成
- 核開発問題を巡り対立していたイランとは15年7月、英仏独中露の5カ国と共に包括的共同行動計画を結んだ。イランは核兵器取得につながる核計画の規模縮小に応じ、米国など主要国は制裁解除で応えた。キューバとの国交回復も果たした。
- 就任4カ月前に発生したリーマン・ショックで米経済は深刻な金融危機に陥っていたが、米史上最大規模の景気対策に着手。
- イラク駐留米軍撤退という最大の公約を果たしながら、過激派組織「イスラム国」(IS)の増勢を許し、2014年に再度イラクへの軍事介入を余儀なくされた
- 大統領は現実的でなく脇が甘い」など批判的な意見

図5: 被験者Aによる「オバマ」に関するまとめ(ベースラインシステム利用)

- TPPを政権の「遺産(レガシー)」にしたい
- 「雇用を生み、環境にやさしい経済を打ち立てるため、革新の精神を育てないといけない」と強調
- プラハ演説で「核兵器なき世界」を提唱
- 広島市の平和記念公園で改めて、「核兵器なき世界」の理想を追い求める決心を強調
- 停滞する核軍縮機運を再び高める
- 任期の最後に「レガシー(政治的な遺産)」を築く狙いがある
- 国内では医療保険制度改革(オバマケア)、リーマン・ショックで落ち込んだ米国経済の回復に取り組んだことを挙げた
- 経済、軍事分野で台頭する中国を念頭にアジア太平洋地域を重視する「リバランス(再均衡)」政策を進め、ASEANとの連携を重要な柱の一つとしている
- 過激派組織「イスラム国」(IS)や国際テロ組織アルカイダの脅威が続く状況で大統領選の年を迎えたため、一般教書演説で、「米国民を守り、テロ組織と戦うことが最優先事項だ」と明言
- 弱腰政権と批判された

図6: 被験者Aによる「オバマ」に関するまとめ(提案システム利用)