

コンテンツ構成要素の順序特性に基づく比較ニュース検索方式

北山 大輔[†] 角谷 和俊[†]

[†] 兵庫県立大学大学院環境人間学研究科 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

E-mail: †ne07p001@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

あらまし 現在、TV や新聞、インターネットなどを通して映像やテキストのニュースコンテンツが配信されている。異なるメディアで提供されるニュース記事は相補的な関係にあると考えられ、同時に提示することで、より理解を深めることが可能であると考えられる。また、メディアの違いにより、同じ内容であっても、異なる使われ方をしている場合がある。例えば、あるメディアでは今後のこととして述べているのに対し、他方のメディアでは主要な出来事として述べているといった具合である。このような、異メディアのニュースを比較して閲覧することで、そのニューストピックの細部まで理解することが可能となると考えられる。しかしながら、その相補的な関係に当たるニュースは、キーワードの出現構造が異なるだけであり、単に内容が同じニュースと区別することが困難である。そこで本稿では、ニュースアーカイブに対し、現在閲覧中のニュースコンテンツのメディア構成要素の順序特性を基に質問生成を行い、異なるメディアに対し構造的に比較関係に当たるニュースを検索する手法を提案する。またその中で、順序付質問を定義し、プロトタイプシステムの構築を行い、比較ニュース検索の質問生成結果および検索結果の確認を行う。キーワード 情報検索、情報統合、ニュースアーカイブ、質問生成

A Comparative News Retrieval Method using Content Elements Order

Daisuke KITAYAMA[†] and Kazutoshi SUMIYA[†]

[†] Graduate School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: †ne07p001@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

Abstract Video and text-news content have been broadcast on TV, newspapers, and the Internet. We consider that there is a complementary relation between news articles of different media. It is possible to be promoted better understanding by comparing these articles. However, it is difficult to distinguish the complementary article from the same content article. Because the appearance structure of the keyword is only different. We propose a comparative query generation method using order of content element. It is composed of two parts. In the first part, analyzing order of news element from user's viewing news. In the second part, generating queries using order feature of news element for retrieving comparison news.

Key words Information retrieval, Information integration, News archives, Query generation

1. はじめに

ニュースによる情報伝達は TV や新聞のみならずインターネットにおいても一般的となってきた。近年、映像ニュースはインターネット上でも FNN-NEWS.COM [2], TBS News i[15] といった各報道局により映像ニュースサイトが公開されてきている。また、テキストニュースもインターネット上の各社のウェブサイト (Sankei Web [13], MSN Mainichi Interactive [9] など) で公開されている。GoogleNews [3] のように、映像やテキストといった異メディアのニュースコンテンツを統合して提

示するサービスも行われている。しかしながら、これらのニュースサイトでは、続報や同一記事といった観点からの関連付けしか行われておらず、ユーザが、ニュースコンテンツを理解するのに十分であるとはいえない。

一般に、より深い理解を得る場合、比較するという行動を取ると考えられる。すなわち、ニュースの理解を助けるためには、ニュース中の人物や出来事、またその動作についての比較ができるコンテンツを検索して提示することが必要であると考えられる。このようなニュースは、同内容が含まれるにもかかわらず、コンテンツの構成が異なる記事であると考えられる。例

例えば、あるニュースでは主題として言われている内容が、他のニュースでは補足程度の扱いになっているという具合である。また、メディアによってニュースコンテンツの構成が異なるため、コンテンツがどのように構成されているかを考慮した検索を行う必要がある。そこで本研究では、現在ユーザが閲覧しているニュースコンテンツに対し、内容が類似しているが、その構成が変化しているニュースコンテンツを自動的に検索する仕組みを提案する。

以下、2 節において研究の概要と関連研究について述べ、3 節ではニュースコンテンツ構成要素の順序特性について説明し、4 節で順序に基づく比較ニュース検索を述べる。5 節でプロトタイプについて述べ、最後に 6 節でまとめと今後の課題について述べる。

2. 研究の概要と関連研究

2.1 研究の概要

本稿では、ニュースメディアごとのコンテンツ構成要素の順序特性を考慮した比較ニュース検索について述べる。ニュース中より、対象となっている物事として名詞、その動作の表現として動詞を抽出し、コンテンツ構成要素の順序関係を付与した質問を生成することで、ユーザが現在閲覧しているニュースコンテンツの一部の内容が含まれるが、構造が変化しているようなニュースを検索する。このとき、構造が変化しているようなニュースというのは、メディアのコンテンツ構成要素の順序が異なるニュースのことである。

コンテンツ構成要素の順序特性とは、メディアによって異なる、文の出現順序のことである。映像ニュースでは、短時間で内容を的確に伝えるために出来事を時系列的に出現させる傾向にある。例えば、出現順が先であれば時間的に前、出現順が後であれば時間的に後という役割を与えられていると捕らえることが可能である。テキストニュースでは、読み返しや斜め読みなどさまざまな読み方に対応するために、出来事をニュースの理解に必要な順に出現させる傾向にある。例えば、出現順が先であればニュースの主となる内容、出現順が後であれば補足的な内容という役割が与えられていると捕らえることが可能である。

我々は、コンテンツ構成要素の順序特性を利用することで、同一内容の位置づけが異なるような比較ニュースを検索することが可能であると考えた。具体的には、記事から抽出したキーワードの順序関係が異なるような、異メディアの記事の検索を行うことで、抽出したキーワードの記事内での位置づけが異なる記事を得ることである。このようなクロスメディアの比較ニュースを得るために、順序関係を扱える質問を定義し、一つの記事から自動的に順序関係を考慮した質問を生成し、検索する手法を提案する。この時、対象に着目した比較ニュースと、動作に着目した比較ニュースの 2 種類の視点の検索質問を生成することで、比較ニュース検索を実現する。図 1 は本手法の概念図である。本手法の手順は以下のとおりである。

コンテンツ構成要素の順序関係の判定 ニュース記事内のキーワードの順序関係は、複数のキーワードが複数個出現するため、

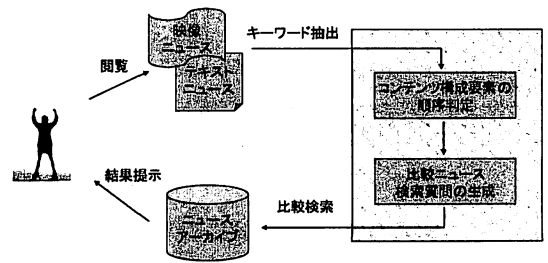


図 1 比較ニュース検索の概念図

一部の順序関係が定まらない、半順序関係であると考えられる。本手法では、構成要素である文単位で順序関係を捕らえ、2 キーワード間の順序関係を判定する。判定を行う基準としては、片方向の順序関係の出現頻度を用い、優位に出現する順序関係を 2 キーワード間の順序として判定を行う。その結果、 $a < b$, $a \circ b$, $a > b$ という、“a が b よりも先”、“a と b の順序は同値”、“a より b の順序が後”の関係を判定する。

順序付質問による比較ニュース検索 クロスメディア比較質問では、通常の検索質問で用いる論理式に加え、順序関係を満たすことを条件として付与した質問の生成を行う。さらに、対象としての名詞を中心に比較を行う質問と、対象の起こした動作である動詞を中心に比較を行う質問の 2 種類を生成することで、2 つの観点から同一の出来事に関する比較を可能とする質問の生成を行う。質問の生成は、抽出されたキーワード集合に、順序関係を付与した状態で、論理式を生成することで行う。

このことにより、ユーザはニュースコンテンツを閲覧するだけで、自動的にニュース内容の位置づけが異なる異メディアコンテンツを得ることが可能となる。

2.2 関連研究

現在提供されているニュースサイトとして McKeown ら [8] の Newsblaster [10] や Radev ら [12] の NewsInEssence [11] や GoogleNews があげられる。これらのニュースサイトは主として、そのトピックを簡潔に理解するための続報記事の集約・要約が目的である。また、ニューストピックを俯瞰的に見る研究として渡邊ら [20] や井手ら [5]、吉岡ら [21] の研究があげられる。これらの研究は、あるニュースを多面的に見るという点で本研究と類似している。しかし本手法は、メディアによる構成の差を比較するために多面的に見るものであり目的が異なる。

キーワードの出現順序に着目して検索を行う研究として、大島ら [23] や山本ら [22] の研究があげられる。大島らは「～や～」といった並列助詞に着目し、検索エンジンのインデックス中から並列助詞の前後に出現する語句を抽出することで、同位語のクイックマイニングを提案している。山本らは、与えられた命題の一部分に対しての真偽を評価するために命題を分割し、一部分が異なり、かつ残りの命題の順序を維持した文章を Web 中から検索することで、命題の真偽指標を提示する「ほんと？サーチ」というシステムを開発している。これらの研究では、特定のキーワード出現順序を用い、その順序に合致するかで関係を判定しているのに対し、本研究では、動的に順序関係

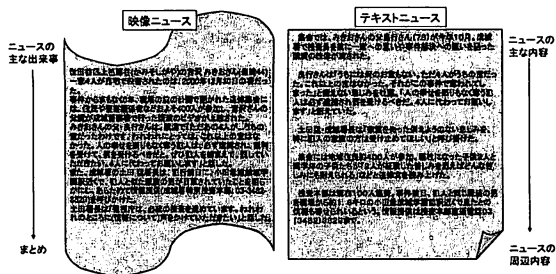


図2 コンテンツ構成要素の順序特性

を抽出することで、記事の比較に適切な順序を自動的に生成した上で検索を行うという点で手法が異なる。

あるコンテンツから検索質問を生成する研究として Henzinger ら [4] や馬ら [18] の研究があげられる。Henzinger らはニュース映像から自動的に質問を生成し、その内容に類似した Web ページを検索する手法を提案している。馬らはテレビ番組から自動的に、内容を幅広くカバーするための質問や掘り下げるための質問など、数種類の質問を生成しテレビ番組の内容を補完できる Web ページを検索する手法を提案している。本研究は、内容の補完や補足を行うのではなく、あるニュースを別のニュースと比較することで内容の理解を深めるということを目的としている点で異なる。

3. コンテンツ構成要素の順序特性と比較関係

3.1 コンテンツ構成要素の順序特性

我々は、同じ内容のニュースコンテンツでも、メディアの違いによって構造に差が現れると考えた。映像ニュースでは、時間軸に沿うため、コンテンツの内容は一方的に伝達されるのに対し、テキストニュースでは、文章表現であるため、読み返しや斜め読みなど多様な伝達に対応する必要がある。このため、この構造の差は、コンテンツを構成する要素である文の順序に現れていると考えられる。

映像ニュースの内容の順序は、短時間で内容を的確に伝えるために時系列的に並びやすく、「発生 (Lead-in)、現状 (Body)、まとめ (Standupper)」という順序で述べられることが多い [17]。すなわち、より先に出現する文は、過去の出来事について述べている可能性が高く、逆により後に出現する文は、今後の方向性やまとめなおしといった、ニュース全体の締めくくりを述べている可能性が高い。一方、テキストニュースの内容の順序は一般に、「概要、詳細、補足」というように、ニュースを理解するために必要な事象順 (逆ピラミッド型) に述べられることが多い [16]。すなわち、より先に出てくる文によって、ニュースの主となる内容を述べている可能性が高く、より後に出現する文によって、ニュースの補足的な内容について述べている可能性が高い。図 2 にコンテンツ構成要素の順序特性の例を示す。映像では、出来事の発生から述べられ、最後に今後の方針が述べられていることがわかる。テキストでは、集会で行われた講演というニュースの主たる内容を先に述べ、文が進むにつれて、補足的な周辺内容が述べられていることがわかる。我々はこの

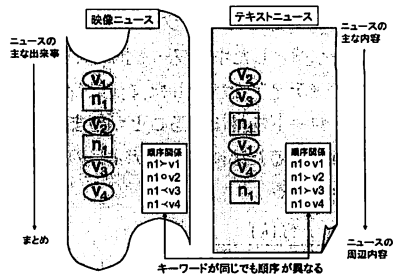


図 3 順序特性に基づく比較関係

ように、メディアごとにコンテンツの内容を述べる出現順が異なることを予備実験により確認している [19].

3.2 コンテンツ構成要素の順序特性に基づく比較関係

比較関係を抽出する目的は、ある記事と、同じ内容が含まれるがその位置づけが異なる記事の両者を見比べることで、より深い理解を得ることにある。このような、同じ内容であるがその位置づけが異なるようなニュースは、メディアが異なる場合に抽出可能であると考えられる。すなわち、映像ニュースで主として扱われていた内容であれば、テキストニュース中から補足として扱っている記事を抽出することで、その内容の位置づけに関しての比較を行うことが可能である。

比較を行えるニュースは、比較元のニュースに対して、比較基準となるキーワードの順序が相対的に変化していると考えられる。このとき、相対的に順序が変化するというのは、動作集合に対する対象の出現順序、もしくは対象集合に対する動作の出現順序である。言い換えると、動詞集合に対する名詞の出現順序、もしくは名詞集合に対する動詞の出現順序といえる。図3に順序特性に基づく比較関係のモデルを示す。左は映像ニュース、右はテキストニュースである。図中の四角で囲まれた n_1 は名詞であり、丸で囲まれた v_1, v_2, v_3, v_4 は動詞であり、縦軸は出現順序である。このように、出現するキーワードが同じでも、メディアが異なり、順序関係が異なるようなニュースが比較関係にあたると考えられる。例えば、「安倍内閣の年金対策」を扱う映像ニュースであれば、「年金」というキーワードが相対的にはじめの方に出現すると考えられるのに対し、「安倍内閣の社会保険政策」を扱うテキストニュースでは、「年金」というキーワードを補足として相対的に後に出現させると考えられる。この2つのニュースの場合であれば、政策に関するニュースであるため、「審議する」「通過する」といった動詞が共通して出現すると考えられる。すなわち、このような動詞に対する、「年金」という名詞の出現順序の変化具合によって、比較関係を抽出することが可能である。

4. 順序付質問による比較ニュース検索

4.1 質問キーワードの重要度算出

クロスメディア比較質問の生成は、まず閲覧中の記事からキーワードの抽出を行い、次にキーワード間の順序関係の判定、最後に順序付質問を生成することで行う。ここでは、まずキーワード抽出のための重要度算出について説明を行う。

クロスメディアに対応したキーワード抽出を行うためには、互いのメディアを基準としたキーワード抽出を行う必要がある。すなわち、双方のメディアを統合した文書集合の中で特徴的に出現しているキーワードを抽出することが望ましい。このようなキーワードを抽出するために、各メディアの文書頻度を正規化した tf-idf 値を用いる。

$$tf-idf = \frac{tf(k)}{df_v(k) + \alpha \times df_t(k)} \quad (1)$$

式中の k は重要度を算出するキーワードであり、関数 tf によって現在閲覧中の記事に含まれる単語出現頻度、関数 df_v によって映像ニュース中の文書頻度、関数 df_t によってテキストニュース中の文書頻度を算出する。 α はメディアごとの文書数を正規化するための重みである。この式により、各メディアの文書頻度を正規化した tf-idf 値を算出する。

算出した tf-idf 値上位の名詞、及び動詞を用いて質問の生成を行う。このとき、質問生成のために 2 種類の集合を用いる。一つは、tf-idf 値最上位の名詞と、その名詞を含む文中に出現する中で tf-idf 値が最上位の動詞のセットに、tf-idf 値上位の動詞を加えた集合。もう一つは、tf-idf 値最上位の動詞と、その動詞を含む文中に出現する中で tf-idf 値が最上位の名詞のセットに、tf-idf 値上位の名詞を加えた集合。以降前者を対象集合、後者を動作集合と呼ぶ。

4.2 順序関係の判定

閲覧中の質問生成元となるニュース記事や、検索対象の記事中のキーワード間に半順序関係を与えるための判定方式について説明する。記事中から 2 つのキーワードを抽出し、2 キーワード間の順序を判定することで行う。判定は以下の式を用いて行う。

$$order(A, B) = \begin{cases} A < B & (count(a < b) > count(a > b)) \\ A \circ B & (count(a < b) = count(a > b)) \\ A > B & (count(a < b) < count(a > b)) \end{cases} \quad (2)$$

式中の A, B はそれぞれキーワード a の集合、 b の集合を現している。関数 $count(a < b)$ により、記事中の個々のキーワード a の後に出現するキーワード b の個数を算出する。関数 $count(a > b)$ は、逆にキーワード b の後に出現する a の個数を算出する。この時、前後関係の判定は文単位で行い、同一文中に出現しているキーワードの前後関係は判定しない。すなわち、個々のキーワードにおける関係の総和で有意となる順序関係を、キーワード間の関係とする。 $A < B$ は A と B の順序が A が先であることを表し、 $A > B$ は B が先であることを表す。また、 $A \circ B$ は順序関係が同値であることを表している。

4.3 順序付質問

本手法では質問生成の際に、順序関係を扱えるように拡張した論理式を用いる。我々はこの論理式を順序付質問と呼ぶ。順序付質問は、検索結果のコンテンツ内における構成要素出現順序を考慮した論理式であり、以下のように定義する。

$$(a < b) = (a \wedge b) \cap \neg((a > b) \cup (a \circ b)) \quad (3)$$

$$\neg(a < b) = \neg(a \wedge b) \cup ((a > b) \cup (a \circ b)) \quad (4)$$

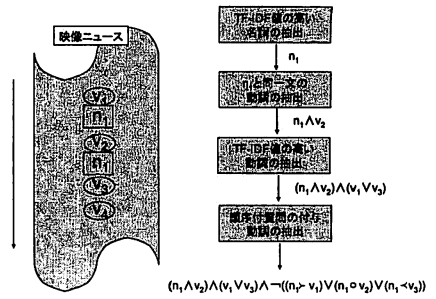


図4 比較質問の生成

すなわち順序付質問は、通常の AND 条件にさらに順序関係を満たす条件を加えたものであり、より条件が厳しい質問となる。逆に順序付質問に NOT 条件が付与された場合には、順序関係の条件を満たさないという条件しかなくなるため、順序関係を満たさなければ、キーワードを含んでいてもかまわないことになる。つまり、通常の NOT 条件より緩やかな質問といえる。言い換えると、AND 条件部分を異なる 3 つの部分集合に分割し、扱えるように拡張した式が順序付質問といえる。

4.4 比較ニュース検索

クロスメディア比較ニュース検索では、対象比較、動作比較の 2 種類の比較質問を生成し、その結果集合を比較検索結果として提示する。まず、対象比較質問の生成について説明し、次に動作比較質問の生成について説明する。対象比較質問の生成は以下の手順で行う。

- (1) 対象集合中の tf-idf 値最上位の名詞 (a) と対応する動詞を AND 条件とした論理式を生成
 - (2) 個々の動詞を OR 条件とした論理式を生成
 - (3) 個々の動詞と名詞 a に関して抽出された順序を用いて順序付論理式を生成
 - (4) 3. で生成された論理式を NOT 条件として、1.2.3. で生成された論理式を AND 条件として論理式を生成
- このようにして生成された質問により、比較軸として抽出した名詞 a と対応する動作を表す動詞を含み、かつ同カテゴリではあるが、名詞 a の位置づけが異なるニュースを検索することが可能となる。すなわち、人物や物事の位置づけが変化したようなニュースを検索することが可能であるといえる。対象比較質問の生成手順を図 4 に示した。

次に、動作比較質問の生成は以下の手順で行う。

- (1) 動作集合中の tf-idf 値最上位の動詞 (b) と対応する名詞を AND 条件とした論理式を生成
 - (2) 個々の名詞を OR 条件とした論理式を生成
 - (3) 個々の名詞と動詞 b に関して抽出された順序を用いて順序付論理式を生成
 - (4) 3. で生成された論理式を NOT 条件として、1.2.3. で生成された論理式を AND 条件として論理式を生成
- 動作比較質問により、比較軸として抽出した動詞 b とそれに対応する動作主体である名詞を含み、かつ同トピックではあるが、動詞 b の位置づけが異なるニュースを検索することが可能と

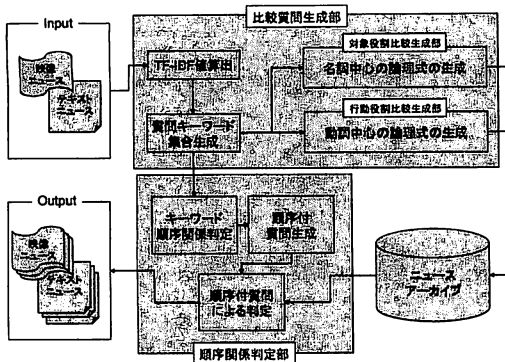


図 5 システム構成

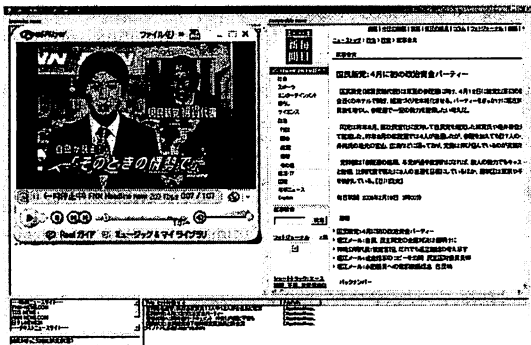


図 6 プロトタイプシステムの画面イメージ

表 1 名詞の tf-idf 値

名詞	tf-idf
綿貫	2.03
きっこう	1.89
拮抗	1.84
新党	1.56
民営	1.47

表 2 動詞の tf-idf 値

動詞	tf-idf
離党	0.99
結成	0.82
つく	0.81
連携	0.76
握る	0.71

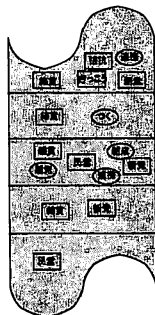


図 7 キーワードの出現順序

order(綿貫,離党) = 綿貫<離党
order(綿貫,結成) = 綿貫<結成
order(綿貫,つく) = 綿貫<つく
order(綿貫,連携) = 綿貫<連携

order(離党,綿貫) = 離党>綿貫
order(離党,きっこう) = 離党>きっこう
order(離党,拮抗) = 離党>拮抗
order(離党,新党) = 離党>新党
order(離党,民営) = 離党>民営

時、直後に「する」が出現する「サ変名詞」は動詞として扱っている。

図 6 にプロトタイプシステムの画面イメージを示す。ユーザは画面左上でニュースコンテンツを閲覧することが可能である。この時、システムは自動的にキーワード抽出を行い、比較質問を生成する。また、生成した質問を用いて自動的にニュースアーカイブに対し検索を行い、検索結果を画面右下のリストボックスに表示する。ユーザは、検索結果を選択することで比較ニュースを得ることができる。また、生成されたクエリを左下のボタン上のテキストボックスに表示するため、どのような条件で検索された結果なのかを知ることが可能である。現在見ているニュースサイトを切り替えたい場合は、左側のリストボックスから選択することで自由にニュースサイト間を行き来することを可能とした。

5.2 自動質問生成の例

作成したプロトタイプを用いた実際の質問生成の例を説明する。質問生成に用いたニュースは、映像ニュースサイト FNN-NEWS.COM の 2007 年 06 月 05 日の「国民新党・綿貫代表、与党が過半数割れた場合の自民党との連携に含み残す」というニュースである。この映像ニュースから抽出した名詞および動詞とその tf-idf 値の一部を表 1, 2 に示す。また、そのキーワードの出現順序および順序関係判定結果を図 7 に示す。この映像ニュースに含まれる文は 5 文であり、図 7 の区切りによって表している。図 7 中では、対象である名詞は四角、動作である動詞は丸で囲んでいる。この結果、対象比較質問のためのキーワード集合として、tf-idf 値最上位の名詞である“綿貫”が抽出される。次に、“綿貫”と同一文に出現する中でもっとも tf-idf 値が高い動詞である“離党”が抽出される。最後に、ニュース中の動詞において tf-idf 値が高い“結成”、“つく”、“連携”が抽出される。図 7 の順序判定結果より、対象比較質問として表

なる。すなわち、動作主体となった人物や出来事の動作に対する位置づけが変化したようなニュースを検索することが可能である。

5. プロトタイプシステム

5.1 システム構成

プロトタイプシステムは大きく分けて、比較質問生成部順、順序関係判定部の 2 つの部分からなる (図 5)。プロトタイプシステムの質問生成は、通常の論理式による検索質問の生成と、結果記事集合中の順序関係基準の生成を行うことで実現している。順序関係判定部では、入力キーワード集合の入力された記事における順序関係の判定を行っている。プロトタイプで用いたニュースアーカイブは、FNN-NEWS.COM, TBS NEWS i の 2 種類の映像ニュースサイト、Sankei Web, MSN Mainichi Interactive の 2 種類のテキストニュースサイトを 1 年 6 ヶ月アーカイブしたものであり、ネットワーク上に配置している。ニュースアーカイブ内の検索は Interstage Shunsaku Data Manager [6] を用いた。検索のインデクスとしては、記事タイトル、記事中の名詞、記事中の動詞、ニュースコンテンツの格納場所を XML で記述したものを用意した。このため、名詞、動詞に分割をした質問を扱うことが可能である。順序関係判定部、質問生成部はともに Visual Studio 2005 の C# により作成した。ニュースコンテンツからの単語抽出には SlothLib [14] を用いた茶筌 [1], [7] による形態素解析を用い抽出した。この

表 3 自動生成される質問の例

対象比較質問	(綿貫 ∧ 離党) ∧ (結成 ∨ つく ∨ 連携) ∧ ¬((綿貫 ∨ 離党) ∨ (綿貫 ∨ 結成) ∨ (綿貫 ∨ つく) ∨ (綿貫 ∨ 連携))
動作比較質問	(離党 ∧ 綿貫) ∧ (きっこう ∨ 拮抗 ∨ 新党 ∨ 民営) ∧ ¬((離党 ∨ 綿貫) ∨ (離党 ∨ きっこう) ∨ (離党 ∨ 拮抗) ∨ (離党 ∨ 新党) ∨ (離党 ∨ 民営))

表 4 検索結果

対象	国民新党：4月に初の政治資金パーティー
比較検索	国民新党：自民離党の下地幹郎衆院議員と新会派
動作	首相指名選挙：郵政造反議員17人中12人が安倍氏に投票
比較検索	国民新党：4月に初の政治資金パーティー 安倍政権へ：同時進行・ドキュメント 仲良し内閣に不安も 国民新党：自民離党の下地幹郎衆院議員と新会派

3の質問が生成される。また、動作比較質問のためのキーワード集合として、tf-idf値最上位の動詞である“離党”が抽出される。次に、“離党”と同一文に出現する中でもっともtf-idf値が高い名詞である“綿貫”が抽出される。最後に、ニュース中の名詞においてtf-idf値が高い“きっこう”、“拮抗”、“新党”、“民営”が抽出される。図7の順序判定結果より、動作比較質問として表3の質問が生成される。

5.3 検索結果の例

生成した検索質問により得られた結果を表4に示す。対象比較質問と動作比較質問の2種類の質問生成を行っているが、双方の検索結果に同じ記事が含まれていた。生成される質問は異なるが、双方の質問とも目的はキーワードの位置づけが異なる記事を検索するという同じ物であるため、適切に検索できている結果であると考えられる。両方で検索されている「国民新党：4月に初の政治資金パーティー」を例として、検索されたコンテンツを説明する。質問生成を行った記事では、“綿貫”という人物を中心に記事が構成されていたが、検索結果の記事では、国民新党を中心に述べられており、“綿貫”は説明のために用いられているに過ぎない。しかし、“離党したが自民党との連携を考えている”といった内容は含まれており、この記事を用いて比較することにより、“綿貫氏個人の考えではなく、党としての考えである”という可能性を読み取ることができる。

6. ま と め

本稿ではコンテンツ構成要素の順序特性を用いた比較質問の生成方式の提案を行った。またその中で、コンテンツに含まれるキーワードの順序関係について定義し、順序付質問の定義を行った。提案方式に基づき、プロトタイプシステムを構築し、質問生成例、及び検索結果例を示した。今後の課題としては、以下のことがあげられる。

- 大規模なニュースデータを用いた有効性の評価実験
- 順序を用いない質問生成との比較実験
- 比較可能な提示インタフェースについての検討

謝 辞

本研究の一部は、平成19年度度富士通 Shunsaku アカデミック支援プログラムによるものです。ここに記して謝意を表すものとします。

文 献

- [1] Asahara, M. and Matsumoto, Y.: Extended Models and Tools for High-performance Part-of-Speech Tagger., *Proc. of The 18th International Conference on Computational Linguistics (COLING 2000)*, pp. 21–27 (2000).
- [2] FNN-NEWS.COM: <http://www.fnn-new.com/>.
- [3] Google news: <http://news.google.com/>.
- [4] Henzinger, M., Chang, B.-W., Milch, B. and Brin, S.: Query-Free News Search., *Proc. of the 12th International World Wide Web Conference(WWW2003)*, pp. 1–10 (2003).
- [5] Ide, I., Mo, H., Katayama, N. and Satoh, S.: Threading News Video Topics., *Proc. of Fifth ACM SIGMM Intl. Workshop on Multimedia Information Retrieval (MIR2003)* (2003).
- [6] Interstage Shunsaku Data Manager: <http://interstage.fujitsu.com/jp/shunsaku/>.
- [7] Japanese Morphological Analysis System ChaSen: <http://chasen.naist.jp/hiki/ChaSen/>.
- [8] Mckeown, K., Barzilay, R., Evans, D., Hatzivassiloglou, V., Klavans, J. L., Nenkova, A., Sable, C., Schiffman, B. and Sigelman, S.: Tracking and summarizing News on a Daily Basis with Columbia's Newsblaster., *Proc. of the Human Language Technology Conference* (2002).
- [9] MSN Mainichi Interactive: <http://www.mainichi-msn.co.jp/>.
- [10] Newsblaster: <http://www1.cs.columbia.edu/nlp/newsblaster/>.
- [11] NewsInEssence: <http://www.newsinesence.com/nie.cgi>.
- [12] Radev, D., Otterbacher, J., Winkel, A. and Blair-Goldensohn, S.: NewsInEssence: summarizing online news topics., *Communications of the ACM*, Vol. 48, pp. 95 – 98 (2005).
- [13] Sankei Web: <http://www.sankei.co.jp/>.
- [14] SlothLib: <http://www.dl.kuis.kyoto-u.ac.jp/SlothLibWiki/>.
- [15] TBS NEWS i: <http://news.tbs.co.jp/>.
- [16] ウィキニュース:スタイルマニュアル: <http://ja.wikinews.org/wiki/ウィキニュース:スタイルマニュアル>.
- [17] ニュースの分析: <http://akasaka.cool.ne.jp/kakeru3/bs3.html>.
- [18] 馬強, 田中克己: 話題構造に基づく放送と Web コンテンツの統合のための検索機構., *情報処理学会論文誌*, Vol. 45, pp. 18–36 (2004).
- [19] 北山大輔, 角谷和俊: ニュースアーカイブのためのコンテンツ構成順序を用いた比較ニュース検索., *電子情報通信学会第18回データ工学ワークショップ (DEWS'07) 論文集*, Vol. A9-4 (2007).
- [20] 渡邊拓也, 大野成義, 太田学, 片山薫, 石川博: 差異に注目した複数文書融合手法., *電子情報通信学会第16回データ工学ワークショップ (DEWS'05) 論文集* (2005).
- [21] 吉岡由智, 湯本高行, 田中克己: ニュースの視点の抽出によるマルチメディアニュースアーカイブの利用., *情報処理学会研究報告*, Vol. 2005-DBS-137(II), pp. 415–420 (2005).
- [22] 山本祐輔, 手塚太郎, Jatowt, A., 田中克己: ほんと?サーチ: 検索結果の集約とページ生成時間分布解析による Web 情報の信用度評価., *電子情報通信学会第18回データ工学ワークショップ (DEWS'07) 論文集*, Vol. C3-3 (2007).
- [23] 大島裕明, 小山聡, 田中克己: Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見., *情報処理学会論文誌: データベース*, Vol. 47, No. SIG19, pp. 98–112 (2006).