

F_037

継続性が存在するニュースのためのニュース記事作成支援システムの試作 A News Input Edit Support System for Continuous and Expansive News

伊藤 正都[†] 大園 忠親[†] 新谷 虎松[†]
Masato Ito, Tadachika Ozono, Toramatsu Shintani

1はじめに

近年、ニュース記事に対するリアルタイム性の要求が高くなっている。また、通信技術が発展、およびグローバル化社会によって、ニュースとなる事象が増加しており、ニュース記事の配信数が増加している。しかしながら、ニュース記事の作成に掛けられるコストは、ニュース記事数の増加数と比例しているとは言い難い。結果として、一人のニュース記者が作成するニュース記事数は増加している。

このような状況において、より質の高いニュース記事の作成、および多くのニュース記事の作成には、ニュース記事自体を作成することを、より効率的にする、若しくはより多くの時間をかける必要があると考えられる。つまり、ニュース記事作成時の入力、および校正に対する作業時間を減らす必要があると考えられる。

本稿では、ニュース記事配信に対して、ニュース記事自体の質、および量の向上には、ニュース記事自体の作成をより効率的にする必要があると考え、ニュース・バリュー論における、ニュースの発展性、および継続性に基づき、ニュース記事の作成をより効率的にするニュース記事作成支援システムの提案を行う。具体的に、ニュース記事を作成する際に、過去のニュース記事から入力中のニュース記事に対し、関連性の高い語を列挙する。そして、ニュース記者に対し列挙された語からの入力を促すことで、ニュース記者がニュース記事を作成する際の入力コストを下げるニュース作成支援手法を提案する。

本稿の構成を以下に示す。第2章では、ニュースの継続性に基づく記事作成支援に関する研究を示す。第3章では、ニュース記事作成支援システムの概要を述べる。最後に第4章で本稿をまとめる。

2 ニュースの継続性に基づく記事作成支援

ニュース・バリュー論の先駆者として知られる Galtung[1] らは、ニュース・バリューの構成要素において次の12項目を挙げている。周期性、強度、明確、意義があること、調和性、意外性、継続性、構成、大国であること、エリートであること、出来事を人格的に語れること、および負の内容をもつていていることである。また、タックマン[2] は次の5項目を挙げている。

硬いニュース、柔らかいニュース、スポットニュース、展開中のニュース、および継続ニュースである。

Galtung、およびタックマンの主張によりニュースを構成する要素として、ニュース記事の継続性、および展開性が含まれることがわかる。

本研究では、Galtung、およびタックマンのニュース・バリュー論に基づきニュース記事の展開性、および継続性に着目する。ニュースの発展性、および継続性とは、あるニュース記事Bに対して展開、および継続において元となるニュース記事Aが存在する性質を示す。つまり、ある事象に対して、ニュース記事として配信される事象の経緯を展開性、および継続性と言う。あるニュース記事に対して展開、および継続が存在する場合、ニュース記事内においては、元となるニュース記事内に使用されている語が出現する確率が高いと考える。このことから、ニュース記事作成時に、配信時間の新しいニュース記事において、同一の語が出現した文章から出現する名詞、および未知語を抽出し、それらの語をニュース記者に提示することでニュース記事作成を支援する。

予測入力方式は、かな漢字変換ソフトウェア、および携帯電話の入力方式として広く利用されている。予測入力方式では、ユーザが過去に入力した単語、および文章に基づいて学習が行われる。これは、ユーザが過去に入力した単語、および文章に大きく依存をすることを示している。現在、多くのかな漢字変換ソフトウェア、およびカナ漢字変換辞書はユーザ毎に個別に使用される場合が多く、かな漢字変換辞書を共有しない。具体的に、ユーザXが単語 α を利用したというデータを、ユーザYが利用することができない場合があることを示している。

一般的に、かな漢字変換ソフトウェアを用いた予測入力方式では、複数ユーザにおいて継続性、および発展性が存在するニュース記事作成に関して、ニュース記者の入力履歴、および実際のニュース記事を有効に活用できない。

3 ニュース記事作成支援システム

3.1 システム概要

本システムのシステム構成図は図1である。ニュース記者は本システムにおいてニュース記事を入力する。入力されたニュース記事は Cabocha[3] によって形態素

[†]名古屋工業大学 天学院 情報工学専攻

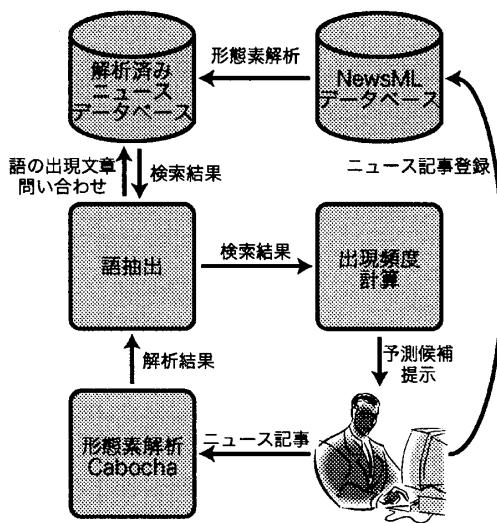


図1: システム構成図

解析を行う。形態素解析の結果から語を抽出し、関連性の高いニュース記事を解析済みニュースデータベースから検索する。検索結果から、語の出現頻度を計算し、ニュース記者に対して関連性の高い語である予測入力候補語の提示を行う。語の抽出、および出現頻度計算に関しては後述する。作成されたニュース記事は、XML形式のニュース配信フォーマットであるNewsMLとしてNewsMLデータベースに登録される。また、作成されたニュース記事に対して形態素解析を行い、解析済みニュースデータベースに登録をする。

3.2 語抽出

本システムでは、Cabochaによって形態素解析、および固有表現抽出を行うものとする。抽出された固有表現、一般的に名詞、および未知語をクエリワードとし、形態素解析済みニュースデータベースから抽出された語を含むニュース記事を抽出する。抽出された形態素解析済みの文章から出現する名詞、および未知語を集計する。この際に未知語を検索対象にするのは、最新の出来事を基に作成されるニュース記事の性質上、Cabochaのに用いられる辞書に登録されていない単語が出現することがあり、それらが未知語として出力されるためである。また、出力された未知語が、ニュース記事において重要な語となる可能性があるからである。

3.3 日付指向性抽出

本システムでは、配信日時、およびニュース記事本文からニュース記事の示している日付を抽出するものとする。これは、ニュース記事において、行事の予定、および結果を配信する場合があり、関連性の高いニュース記事は語出現頻度計算時に考慮に入れる必要があると考えられるからである。事前に、日付指向性を抽出しておくことで、語出現頻度計算時に大きな影響を与えると考えられる。

3.4 語出現頻度計算

ニュース記事の配信時間を参考にし、語抽出の過程において集計された名詞、および未知語を新しい記事での出現頻度が高いほど影響力が高くなるように順位付けする。これは、ニュース記事が展開、および継続の性質を持っており、同一の語が出現するニュース記事に関して、配信日時がより最近のニュース記事ほどニュース記者が編集中のニュース記事に対して関連性が高いと考えられるためである。次に語出現頻度計算の計算式を示す。

$$F_W = \frac{\sum_{t=0}^{30} \{N_{Wt}(d)^t\}}{M} + C \quad (1)$$

式(1)において、 t はニュース記事が配信されてからの時間(日)、 N_{Wt} は t 日前のニュース記事に出現したある語 W の回数、 d は過去のニュース記事からの影響減衰率、 M は30日以内に配信されたニュース記事総数、 C は日付指向性における補正值、 F_W はある語 W が30日以内に出現した頻度を表す。 F_W を基に、語出現頻度を順位付けする。

3.5 Web ブラウザ上の動作

Web アプリケーションとして、本システムを実装することでネットワークを経由してニュース記者はニュース記事の投稿、および編集が可能となる。また、ニュース記者に対して新しいアプリケーションのインストール作業等、煩わせること無くニュース記者が普段使用している環境のまま利用可能である。

4 おわりに

本稿では、ニュース記事の継続性、および展開性における同一語の出現確率の高さに着目し、予測入力候補語を示すことでWebブラウザ上で動作するニュース記事作成支援システムを提案した。本システムにより、ニュース記者はニュース記事をより容易に作成できる。具体的に、ニュース記事入力時の入力コストを減少させる。また、ニュース作成時の校正作業を軽減させ、作成時間を短縮させる効果が考えられる。

今後の課題として、語出現頻度計算式(1)における、影響減衰率 d 、および補正值 C の最適値を求めることが挙げられる。

参考文献

- [1] Galtung.J, Ruge.H, The Structure of Foreign News, Journal of Peace Research vol.1, 1965
- [2] タックマン・G, ニュース社会学, 三嶺書房, 1991
- [3] 工藤拓, 松本裕治, チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.2002, No.043, 2002