

マルチコンテンツによる「ニュースの言葉」 自動生成システムの研究

小林透^{†1} 柴田和樹^{†2} 中山僚^{†3}

もっともポピュラーなニュース媒体である新聞には、その日の出来事に関する専門用語を解説する「ニュースの言葉」というコラムがある。このコラムは、普段聞きなれない専門用語を理解する上で有益である。しかしながら、取り上げられる専門用語は、それぞれの新聞社が独自に選択したものであり、かならずしも読者が知りたい専門用語が取り上げられるとは限らない。また、解説内容もそれぞれの新聞社が独自に編集しており、内容に偏りが生じる場合がある。そのため、さまざまな背景知識を持つ読者すべてにとって必ずしも理解しやすいとは言えない。そこで、本研究では、その日のニュースから複数の専門用語を自動抽出する。そして、その専門用語を理解する上で有用な一般概要情報、ソーシャルビデオ情報、ツイート情報、関連プロダクト情報と言った多面的な関連情報を自動収集する。さらに、収集された情報をマルチスクリーン環境にマルチモーダルユーザインターフェースにより表示する。これにより、読者自らがマルチコンテンツな「ニュースの言葉」を自動生成し、それをマルチスクリーン環境にアクセスビリティの高いマルチモーダルユーザインターフェースで表示するシステムを実現する。

1. はじめに

もっともポピュラーなニュース媒体である紙ベースの新聞には、その日の出来事に関する専門用語を解説する「ニュースの言葉」というコラムがある。このコラムは、普段聞きなれない専門用語を理解する上で有益である。しかしながら、取り上げられる専門用語は、それぞれの新聞社が独自に選択したものであり取り上げられる専門用語の数も限られることから、かならずしも読者が知りたい専門用語が取り上げられるとは限らない。また、解説内容もそれぞれの新聞社が独自に編集しており、新聞社によって内容に偏りが生じる場合がある。そのため、さまざまな背景知識を持つ読者すべてにとって必ずしも理解しやすいとは言えない。

一方、インターネットの普及に伴いマルチメディアに対応した電子新聞やインターネットを利用した情報サービス、外部事業者へコンテンツ提供を行う新聞事業者が増えてきている¹⁾。そのため、スマートフォンを代表とする高機能端末の普及と相まって、特に若い世代ほど紙ベースの新聞を読まずにインターネットのニュースサイトを活用する傾向が高い²⁾。また、新聞事業者が関わるそのようなサービスだけでなく、Wikipedia や YouTube, Twitter に代表されるソーシャル系サービスを活用した情報取得も一般的になってきている。

そこで、本研究では、ネット上で配信されているその日のニュースから複数の専門用語を自動抽出する。そして、その専門用語を理解する上で有用な一般概要情報、ソーシャルビデオ情報、ツイート情報、関連プロダクト情報と言

った多面的な関連情報（マルチコンテンツ）を自動収集する。さらに、収集されたマルチコンテンツをマルチスクリーン環境³⁾に、マルチモーダルユーザインターフェース⁴⁾により表示する。システム化に当たっては、SOA (Service Oriented Architecture) のコンセプトに従い最新の Web 技術、Web サービスの活用が可能な拡張性と柔軟性を持つシステムアーキテクチャとする。

これにより、読者自らがマルチコンテンツな「ニュースの言葉」を自動生成する。そして、読者自らがそれをマルチスクリーン環境上にマルチモーダルユーザインターフェースを介して表示するシステムを実現する。これは、異なる知識やアクセスビリティを持つ読者に対して、ニュースという新規性の高い情報に含まれる未知の専門用語に対する理解性を高めることにつながる。

以下、2 章で関連研究を説明し、3 章で「ニュースの言葉」自動生成システムに求められている要求条件を定義し、それらの要求条件を具現化するシステム構成方法を説明する。4 章で開発したプロトタイプシステムを説明し、5 章でプロトタイプシステムを用いた評価実験結果を示す。6 章でその考察を述べ 7 章でまとめる。

2. 関連研究

電子新聞やネット上のニュースに関する研究としては、情報フィルタリング⁵⁾に関する研究がこれまで多くなされている。これは、読者の興味や嗜好に合わせて情報を篩にかけ、読者が必要とするニュース記事のみを提供したり⁶⁾、読者に提示する順番を変えたりする⁷⁾というものである。その他の関連研究としては、ネット上のニュース記事のダイジェストを自動的に生成することで、情報全体の俯瞰やエッセンスの把握に活用しようという研究⁸⁾がある。しかし、これらの研究は、ニュース記事全体を対象にフィルタリングしたり、ニュースのサマリーを作成したりするものであり、ニュース記事の中から専門用語を抽出して、その関

†1 長崎大学大学院工学研究科
Graduate School of Engineering, Nagasaki University

†2 長崎大学工学部
School of Engineering, Nagasaki University

†3 長崎大学工学部、現長崎県警察
School of Engineering, Nagasaki University. Currently, Nagasaki Prefectural Police

連情報を収集する本研究とは目的が異なる。

一方、専門用語の説明箇所をネット上から収集し、辞典や用語集の自動生成につなげようという研究がある⁹⁾¹⁰⁾。これらの研究は、HTML構文解析やテキスト解析によりネット上から対象とする専門用語の説明と思われる箇所を特定し、それらを収集・整理するというものである。専門用語の理解を助けるという点では、本研究と同じであるが、本研究は、ニュース記事に含まれる専門用語を対象としていること、収集対象の情報はテキスト情報だけではなく、映像情報を含むマルチコンテンツであることが異なる。

3. 「ニュースの言葉」自動生成システム

3.1 要求条件の整理

提案システムは、既存の紙ベースの新聞での「ニュースの言葉」と同等の手軽さで利用ができ、それに加えて新聞の場合のデメリットを解決し、かつ最新のWeb技術を活用したユーザフレンドリな操作性が求められる。これらの読者の使い勝手に関わる具体的な要件は、以下の通りである。

要件1: 受動的な情報取得が可能なこと

要件2: 多面的な関連情報の取得が可能なこと

要件3: アクセシビリティの高いユーザインターフェースの利用が可能なこと

また、近年Web技術やWebサービス技術の発展は目覚ましく、すべてを自ら構築するのではなく、外部の優れた技術やサービスを取り込んでいくことが時間とともにシステムを陳腐化させないことにつながる。そこで、システム構築に関する要件として、以下を取り上げる。

要件4: 拡張性や柔軟性の高いシステムとすること

3.2 処理フロー

提案システムの具体的な処理の流れ(図1)は以下の通りである。図1の中で、□は、システムが実行する処理、◇は読者が行う処理を表す。

①Webニュースサイトからニュース一覧を取得し、読者に提示

一般的なWebニュースサイトでは、最新ニュース、経済、エンターテインメント、スポーツなどニュースのジャンル毎にタブを切り替えることで、ジャンル毎のニュースタイトルを表示する形態が多い。本システムでも、これに準拠し、読者がタブを切り替えることで、ジャンル毎のニュースタイトルの一覧が表示される形態とした。読者は、提示されたニュース一覧の中から気になるニュースタイトルを一つ選択する。これにより、次の処理が開始される。

②Webニュースサイトから選択されたニュースの本文のみを抽出し、読者に提示

一般的なWebニュースサイトでは、ニュース本文の他に、広告や関連ページへのリンク情報などが同一ページに掲載されるケースが多い。そのため、対象となるページの掲載情報を解析することでニュース本文のみを抽出し、読者に

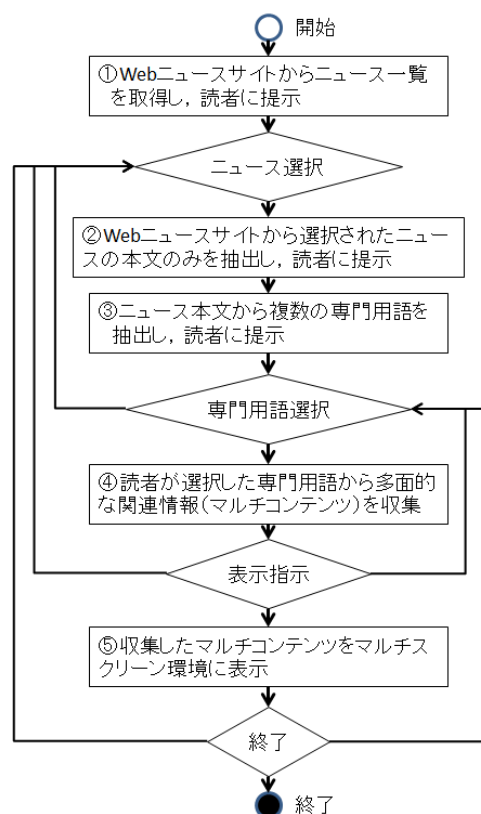


図1 処理フロー

Fig. 1 Process Flow

提示する。

③ニュース本文から複数の専門用語を抽出し、読者に提示

抽出されたニュース本文の中で、読者がさらに知りたいと思われる専門用語を複数抽出し、それらを読者に提示する。これらの抽出された専門用語が「ニュースの言葉」の候補となる。読者は、提示された専門用語の中から気になる専門用語を一つ選択する。これにより、次の処理が開始される。専門用語を選択せず、別のニュースタイトルを選択することも可能である。

④読者が選択した専門用語から多面的な関連情報(マルチコンテンツ)を収集

選択された専門用語に関する一般概要情報、ソーシャルビデオ情報、ツイート情報、関連プロダクト情報を多面的な関連情報としてネット上から収集する。読者は、収集された情報のマルチスクリーン環境での表示方法を指示する。これにより、次の処理が開始される。表示指示を行わず、別の専門用語を選択したり、別のニュースタイトルを選択したりすることも可能である。

⑤収集したマルチコンテンツをマルチスクリーン環境に表示

読者の指示に従い収集したマルチコンテンツをマルチスクリーン環境に分散表示する。その後、本システムを終了させない限り新たなニュースタイトルの選択や専門用語の

選択を繰り返すことが可能である。

図1に示した通り、読者が行うのは、提示された情報の選択と表示の指示のみである。これにより、読者が指定した専門用語の多面的な関連情報が表示される。しかも、これらの操作は、タブレット端末上でのドラッグ&ドロップ、タブレット端末のモーション操作、音声操作などのマルチモーダルなユーザインターフェースにより可能である。本処理フローにより要件1~3までの実現が可能となる。

3.3 システムアーキテクチャ

本システムは、制御サーバ(Control Server)、読者端末(Reader Device)からなっている(図2)。制御サーバは、3.2節で示した処理フローを読者からの指示に基づき行う。読者端末は、主に読者が操作する操作端末(Operation Device)とソーシャルビデオ情報等を表示する情報表示端末(Display Device)からなっている。本システムアーキテクチャの特徴は、SOAの考え方に則り2つのマッシュアップを実現していることである。一つ目のマッシュアップ対象は、外部の複数のWebサービスである。これは、Open APIを利用して多面的な関連情報を収集するために行っている。具体的な外部のWebサービスは、Yahoo, DBpedia, YouTube, Twitter, Amazonである。二つ目のマッシュアップ対象は、複数の読者端末である。これは、HTTPをベースにしている。これにより、読者端末はWebブラウザの搭載のみを前提とすれば良く、読者の身の回りにある様々な端末を利用した操作、情報表示を可能としている。

以上のように制御サーバをWebサービスと読者端末の間のHub的な装置として位置付け、かつ両者間のインターフェースを規定することで役割分担が明確なアーキテクチャとなる。これにより、今後新しい外部サービスを追加したり、既に利用している外部サービスが変更されたりした際に、制御サーバが担う外部サービスとのインターフェース部分の修正のみで対応可能となる。これは、拡張性や柔軟性の高いシステムという要件4を満足することにつながる。

3.4 システム構成

図3に提案システムのシステム構成を示す。制御サーバには、Server APと操作端末用、情報表示端末用のHTMLリソース(Operation AP, Display AP)を配置した。操作端末、及び情報表示端末は、最初に制御サーバにアクセスすることで、必要なリソースを取得し起動する。また、それぞれのAPには、WebSocket¹¹⁾が実装されており、Server AP-Operation AP間、Server AP-Display AP間の通信は、WebSocketを介して行われる。これにより、3つのAPを連動して稼働させることができる。以下、Server AP, Operation AP, Display APの詳細を説明する。

●Server AP

Server APに実装した個々の機能を3.2節で示した処理フローに従って説明する。

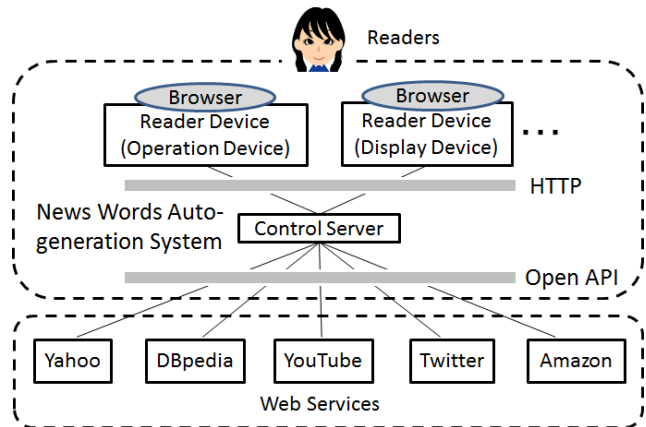


図2 システムアーキテクチャ

Fig. 2 System Architecture

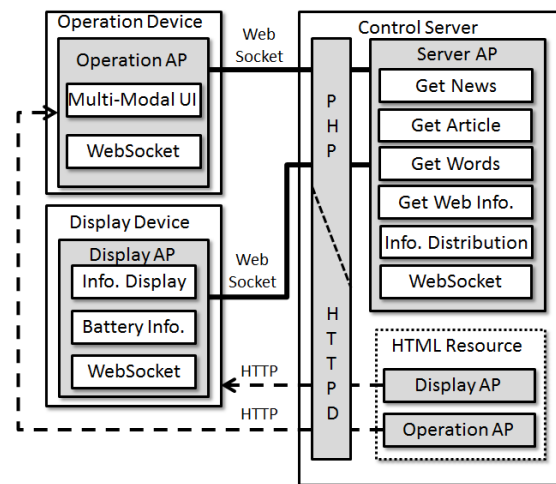


図3 システム構成

Fig. 3 System Configuration

①Get News : Web ニュースサイトからニュース一覧を取得し、読者に提示

Yahoo!デベロッパーネットワークが公開するニュースWeb API¹²⁾を利用して、ニュースタイトル、記事URL等の情報をXMLで取得する。取得した情報は、WebSocketによりOperation APに送信され、ジャンル毎にタブに整理して読者に提示される。

②Get Article : Web ニュースサイトから選択されたニュースの本文のみを抽出し、読者に提示

あるページのニュース記事本文だけを抽出するためには、パーサー機能として、一般的に以下のようなアルゴリズムの実装が必要である。

- HTMLの中のタグや属性情報を利用して記事本文の要素を推測
- 推測した要素内にある広告など、不要な情報を削除

・各種タグ情報を削除し、プレーンテキスト化

これらの機能の高度化が本研究の目的ではないため、フリー・ソフトウェアを活用することとした。具体的には、読者が選択したニュースタイトルの記事本文を示す URL から記事本文を含むページを特定し、このページから Boilerpipe¹³⁾を利用してニュース本文のみを抽出することとした。Boilerpipe は、単語が出現する特徴から確率モデルにより記事本文かどうかを判定する方式¹⁴⁾を取っており、Web API や Java ライブラリが提供されている。また、対象とするページの属性を指定することで、より高精度に本文を抽出することが可能で、属性として“ニュース記事”が指定できる。以上より、要件 4 に適合し、かつニュース記事に特化した抽出機能を持つ Boilerpipe を利用することとした。実装に当たっては、Web API は動作が不安定であったため、Java ライブラリを活用した Java プログラムによりパーサー機能を実装した。

③Get Words：ニュース本文から複数の専門用語を抽出し、読者に提示

読者によって知りたいと思う専門用語は異なるので、本来ならその読者の意向を類推して、ピンポイントで専門用語を抽出して提示することが望ましい。しかし、そのためには、読者毎のプロファイルを収集し高度な分析を行う必要がありコストがかかる。また、仮にこれを行ったとしても、必ずしも高精度な類推を継続的に実現できる保証はない。そこで、本研究では、読者毎に適応させた専門用語の抽出は行わず、Yahoo!デベロッパーネットワークが公開するキーフレーズ抽出 API¹⁵⁾によりすべての読者に同じ専門用語を提示する方式を採用した。この場合、抽出する専門用語の数が少ないと読者が知りたいと思う専門用語と合致しないケースが生じる。そこで、操作端末として想定しているタブレット PC の画面の大きさを考慮し、9つの専門用語を抽出することとした。今後、さらに高度なキーフレーズ抽出機能が利用可能となった場合は、API 部分を改修するだけで適応が可能であり、これは要件 4 に適合する。

④Get Web Info.：読者が選択した専門用語から多面的な関連情報（マルチコンテンツ）を収集

読者が選択した専門用語を検索キーとして DBpedia, YouTube, Twitter, Amazon から公開されている API を利用して関連情報を収集する。Get Web Info.に実装しているのは、API のみなので、仮にこれらの外部サービスが高機能化されて API が変更されても API 部分の改修だけで済む。また、新たな外部サービスを利用するには、その外部サービスの API を追加するだけで良い。これは要件 4 に適合する。

⑤Info. Distribution：収集したマルチコンテンツをマルチスクリーン環境に表示

該当する Yahoo!ニュースの掲載ページや YouTube からのソーシャルビデオ情報を複数の情報表示端末に表示する。YouTube からのソーシャルビデオ情報は最大 4 つの異なる

コンテンツを情報表示端末に表示可能である。

●Operation AP

ブラウザ上でのマルチモーダルユーザインターフェース (Multi-Modal UI) を実装している。具体的には、ドラッグ&ドロップ、モーション、音声の各操作をブラウザ上で動作可能とした。

・ドラッグ&ドロップ操作

マウスクリックとタブレット PC のタッチ操作に対応した処理をそれぞれイベントリスナに登録した。これにより、PC だけでなく、タブレット PC 上でのドラッグ&ドロップを可能とした。

・モーション操作

HTML5 における加速度センサ API である Device Orientation Event Specification¹⁶⁾を利用した。具体的には、X,Y,Z の 3 軸にかかる重力加速度を検出し、タブレット PC の傾け具合による操作を可能とした。

・音声操作

HTML5 における音声認識 API である Web Speech API Specification¹⁷⁾を利用した。あらかじめ決められたコマンドの音声操作だけでなく、ニュースタイトルの一部や専門用語の一部を読み上げるだけで、関連する情報をマルチスクリーン環境に表示できるようにした。これらのマルチモーダル UI 機能は、すべて JavaScript で実装した。

●Display AP

WebSocket を介した制御サーバからの指示により、Yahoo!ニュースの掲載ページや YouTube からのソーシャルビデオ情報をブラウザ上に表示する機能(Info. Display)とバッテリー情報通知機能(Battery Info.)を JavaScript で実装した。情報表示端末は、スマートフォン等バッテリーで駆動しているデバイスが少なくない。そこで、バッテリー残量を Battery Status API¹⁸⁾を利用して取得し、WebSocket により定期的に Operation AP に通知する機能を設けた。

4. プロトタイプシステム

提案システムを、表 1 で示した仕様の制御サーバ、操作端末、情報表示端末を用いてプロトタイプ実装した。図 4 にタブレット PC を想定した操作端末の UI 例を示す。本 UI は、上下方向に画面スクロールをすることで、(a)→(b)→(c)と画面が遷移する。

表 1 プロトタイプ仕様

Table1 Prototype System Specification

Control Server	Operation Device	Display Device
•Linux CentOS 6.4, Apache 2.2.15, PHP 5.3.3, WebSocketServer 2.1.0	•Windows Tablet, Win8 32bit, Chrome 18.0 •iPad Mini, iOS 7.0.2, Safari	•Windows PC, Chrome 18.0 •iPod Touch, iOS 7.0.3, Safari •iPad Mini, iOS 7.0.2, Safari



(a)



(b)



(c)

図 4 操作端末 UI 例

Fig. 4 Operation Device User Interface Image

(a): 上部に「ニュース一覧」が表示され、それぞれ、Topics, Economy 等のタブを選択することで、ニュースのジャンルを切り替えることができる。ここで、一つのニュースタイトルを選択するとそのニュースの記事本文のみが「ニュース記事」として表示される（図 4 の例では、QR コード関係のニュースタイトルを選択している）。そして、その下には、「ニュース記事」に含まれる「ニュースの言葉」が 9 個表示される。

(b): 「ニュースの言葉」を一つ選択するとその概要情報がテキスト情報として表示される（図 4 の例では「QR コード」を選択している）。その下には、関連する YouTube 映像のサムネイルが表示され、最下部の“TVPC1~4”は、利用可能な情報表示端末を表している。YouTube 映像をこれらの情報表示端末に表示するためには、映像のサムネイルを“TVPC1~4”へそれぞれドラック&ドロップするか、タブレット PC を傾げるか、音声により行う。

(c): 選択した「ニュースの言葉」に関するツイート情報、関連書籍の情報が表示される。書籍に関しては、サムネイル画像を選択することで該当書籍の購入ページに遷移することができる。

音声操作に関しては、「ニュース一覧」のタブの切り替えや画面スクロール、オリジナル記事ページへの遷移等の操作系コマンドはもちろん、ニュースタイトルの選択や「ニュースの言葉」の選択においてもその一部を読み上げるだけで可能とした。

図 5 にマルチスクリーン環境での利用例を示す。右下のタブレット PC を操作することで、マルチスクリーン環境にオリジナルのニュース記事を表示させたり、YouTube 映

像を表示させたりすることができる。図 5 に示したように、タブレット PC に提示された「ニュースの言葉」をタップしたり、一部を読み上げたりすることで、一瞬で関連情報が切り替わることが確認できた。これによりユーザが気になる「ニュースの言葉」を次々指定することで、関連情報を次々切り替えることができた。

5. 評価実験

本研究では、3.2 節で示した①~⑤までの処理フローに基づきプロトタイプを開発した。このうち、①はニュースタイトルの一般的な提示形態のため、評価対象から外し、②~⑤までと、処理フロー全体に係るマルチモーダルインターフェースによる操作を評価対象とした。具体的には、以下の処理である。



図 5 マルチスクリーン環境利用例

Fig. 5 Multi-screen Environment Usage Image

- A. ニュースサイトからニュース本文のみを抽出
- B. ニュース本文から専門用語を抽出
- C. マルチコンテンツを収集
- D. マルチスクリーン環境に表示
- E. マルチモーダルインターフェースによる操作

これらの処理は、以下に示す本研究のゴールを達成するために必要なものである。

- F. 異なる知識やアクセシビリティを持つ読者に対して、ニュースという新規性の高い情報に含まれる未知の専門用語に対する理解性の向上

そこで、A~Eの処理、およびFのゴールの達成度を評価するための評価実験を行った。

5.1 実験方法

2014年4月15日、16日に配信されたYahoo!ニュースから無作為に29件の記事を選択し、それらの記事を対象に以下の方法により実験を行った。

- A. ニュースサイトからニュース本文のみを抽出

提案システムによる抽出されたニュース本文と実際のYahoo!ニュースページの記事を目視で比較し正確さを評価する。

- B. ニュース本文から専門用語を抽出

20代~30代の学生を中心とした男女11人(内女性3人)を被験者として、ニュース記事を読んでもらい、その記事の中でさらに知りたいと思う専門用語を1個から最高9個までマークしてもらう。その後、提案システムが抽出した専門用語と比較することで、再現率や適合率を評価する。また、以下のアンケートにより、主観的な評価を行った。「B-1:あなたが選択したニュース記事に関してシステムが

提示した「ニュースの言葉」は、期待通りでしたか？」その後、プロトタイプシステムによる一連の処理の流れを体験してもらった後、C, D, E, Fに関して以下のアンケートにより評価した。

「C-1:あなたが選択した「ニュースの言葉」に関するマルチコンテンツ(DBpedia、YouTube、Twitter、Amazon)の中で、最も有益だと思ったコンテンツはどれですか？」

「C-2:あなたが選択した「ニュースの言葉」に関して複数のコンテンツを提示されることで、より理解は深まりましたか？」

「D-1:あなたが選択した「ニュースの言葉」に関する情報が複数のスクリーンに表示されたことにより、より理解が深まりましたか？」

「E-1:本システムのマルチモーダルインターフェースは、アクセシビリティ向上に役に立つと思いますか？」

「F-1:本システムは、異なる知識やアクセシビリティ能力を持つユーザにとって有効なシステムだと思いますか？」

C-1以外のアンケートは、“とても良い”、“まあまあ良い”、“どちらとも言えない”、“あまり良くない”、“良くない”の5段階で評価してもらった。

5.2 実験結果

B-1からF-1までのアンケート結果を図6に示す。

- A. ニュースサイトからニュース本文のみを抽出

29件のニュースの内、過不足無く完全に抽出できた件数は、22件で、完全抽出率は76%であった。完全に抽出できなかったニュースについては、すべてがニュース記事の一部欠落で、文字数ベースの欠落率の平均は約18%であった。また、ニュース記事本文以外の広告等の不要な情報が混入

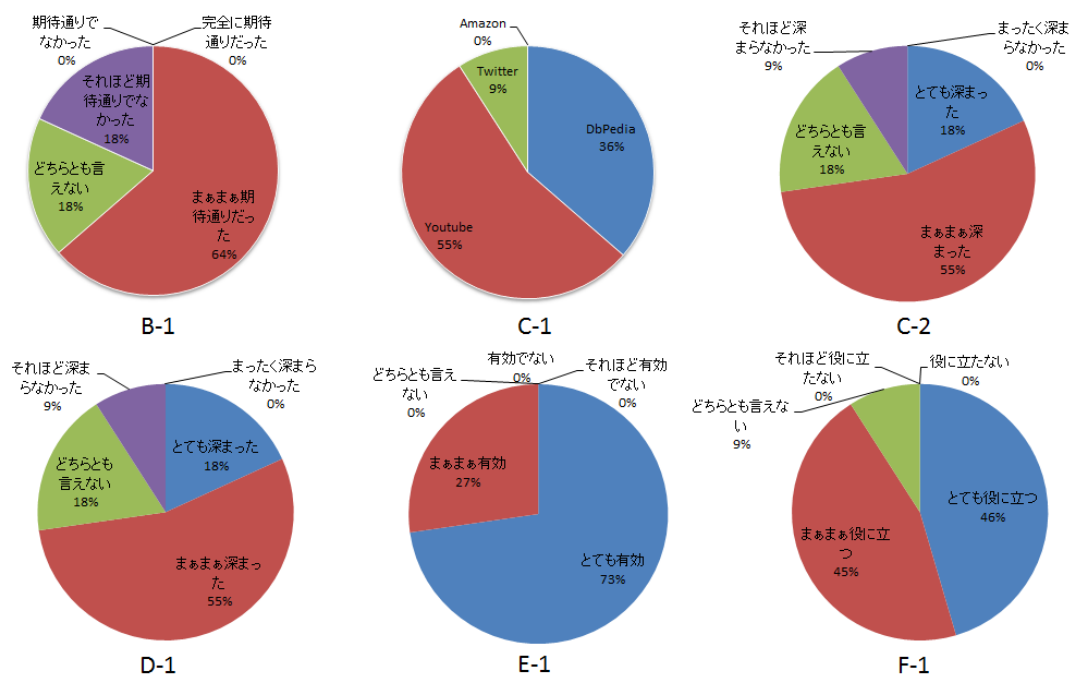


図6 アンケート結果

Fig. 6 Questionnaire Results

するケースは、一件も無かった。

B. ニュース本文から専門用語を抽出

読者がマークした専門用語とシステムが抽出した専門用語の再現率と適合率の平均は、それぞれ、60.5%、23.9%であった。本システムの場合、読者が知りたいと思う専門用語がシステムから提示されたものに含まれていることが重要であるため、適合率よりも再現率が重要となる。この結果から、読者が知りたいと思った専門用語の約6割をシステムが抽出できたことが分かった。また、アンケート結果からは、再現率が6割程度の本結果に対しては、64%の被験者が“まあまあ期待通りだった”という評価を示している。

C. マルチコンテンツを収集

有益なコンテンツとしては、テキスト情報 (DBpedia) よりも映像情報 (YouTube) の方が高い評価を得た。また、マルチコンテンツにより理解が深まったかの質問に対しては、72%の被験者が“とても深まった”“まあまあ深まった”と評価している。

D. マルチスクリーン環境に表示

マルチスクリーンにより理解が深まったかの質問に対しては、被験者の72%が“とても深まった”“まあまあ深まった”と評価している。

E. マルチモーダルインターフェースによる操作

マルチモーダルインターフェースのアクセスビリティ向上に関しては、73%の被験者が“とても有効”という評価をしている。

F. 異なる知識やアクセスビリティを持つ読者に対して、ニュースという新規性の高い情報に含まれる未知の専門用語に対する理解性の向上

被験者の96%が“とても役に立つ”、“まあまあ役に立つ”と評価している。特に46%が“とても役に立つ”と答えている。

6. 考察

5章に示したように、マルチコンテンツ、マルチスクリーン、マルチモーダルインターフェースに関する評価やシステム全体に対する評価は、アンケートによる主観的評価ではあるが高い評価結果が得られた。一方、ニュース本文の過不足ない抽出に関しては、8割弱程度の成功率、読者が期待する専門用語の抽出に関しては、6割程度の再現率であり、改善の余地があることが分かった。以上のことから、本研究のゴールである「異なる知識やアクセスビリティを持つ読者に対して、ニュースという新規性の高い情報に含まれる未知の専門用語に対する理解性の向上を図るシステム」においては、マルチコンテンツ、マルチスクリーン、マルチモーダルインターフェースが有効なアプローチであると言える。また、本文抽出の成功率向上と専門用語の再現率向上により、より読者の満足度を高められること

ができる。

本研究では、単に定義した要件を満足するマルチコンテンツによる「ニュースの言葉」自動生成システムを開発しただけではない。SOA (Service Oriented Architecture) のコンセプトに従い図2で示した最新のWeb技術、Webサービスの活用が可能な拡張性と柔軟性を持つシステムアーキテクチャとそれに基づく実稼働可能なシステム構成を明らかにしたことも成果である。Web技術、Webサービスに関する研究は、日進月歩であり、これらの最新の研究成果を素早く取り込めるかどうかシステムを永続的により良いものとして活用できるかどうかの鍵となる。具体的には、将来的に本文抽出の成功率向上や専門用語抽出の再現率向上が期待できる新たなWebサービスの利用が可能になった場合、API部分の修正だけでそれらに置き換えることができる。また、将来、新たなサービスやコンテンツが登場した際、それらを追加することにより、より読者の理解性を向上させることができる可能性がある。

7. おわりに

本研究では、ネット上のニュース記事に含まれる専門用語を理解する上で有用な一般概要情報、ソーシャルビデオ情報、ツイート情報、関連プロダクト情報と言った多面的な関連情報 (マルチコンテンツ) を「ニュースの言葉」として提示するシステムを開発した。本システムは、収集されたマルチコンテンツをマルチスクリーン環境に、マルチモーダルユーザインターフェースにより表示できることが特徴である。

評価実験の結果、本システムの特徴であるマルチコンテンツ、マルチスクリーン、マルチモーダルインターフェースによる「ニュースの言葉」自動生成に関して高い主観評価が得られた。一方、ニュース本文の過不足ない抽出に関しては、8割弱程度の成功率、読者が期待する専門用語の抽出に関しては、6割程度の再現率であり改善の余地があることが分かった。

本システムは、SOA (Service Oriented Architecture) のコンセプトに従い拡張性と柔軟性を持つシステムアーキテクチャをベースとしている。この特徴を活かし、今後、本文抽出、専門用語抽出のさらなる精度改善、新たなサービス、コンテンツを取り込むことを目指す。これにより、異なる知識やアクセスビリティを持つ読者に対して、ニュースという新規性の高い情報に含まれる未知の専門用語に対するさらなる理解性の向上を目指す。

参考文献

- 1) 日本新聞協会, 新聞・通信各社の電子・電波メディア参入状況, 入手先
<<http://www.pressnet.or.jp/data/media/media01.php>> (2014.05.08).
- 2) リサーチバンク, 新聞に関する調査. 新聞(紙媒体)を「ほぼ毎

- 日読んでいる」人は41%、入手先
<http://research.lifemedia.jp/2013/10/131016_newspaper.html>
(2014.05.08)
- 3) Kobayashi, T., “A Proposal of Smart TV System focused on Findability,” 2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE), pp.507-508 (2013).
- 4) 小林透, “ファインダビリティに着目したマルチモーダルなスクリーンデバイス連携システム”, 映情学技報, Vol.38, No.6, pp.21-24 (2014).
- 5) 森田昌宏, 速水治夫, “情報フィルタリングシステム—情報洪水への処方箋—”, 情報処理, Vol.37, No.8, pp.752-758 (1996).
- 6) 宮本誠一, 小坂雄一, 高谷哲, 土田賢省, 佐藤章, “ネットワークニュースの自動編集システムにおけるキーワード検索の効率化”, 信学技報, COMP95-14, pp.25-32 (1995).
- 7) 曾根直人, 河野仁, 新居和人, 森井昌克, “購読者の趣向を考慮した紙面構成を可能とする電子新聞システムについて”, 信学技報, OFS96-14, pp.25-30 (1996).
- 8) 佐藤円, 佐藤理史, 篠田陽一, “電子ニュースのダイジェスト自動生成”, 情報処理学会論文誌, Vol.36, No.10, pp.2371-2379 (1995).
- 9) 藤井敦, 石川徹也, “World Wide Web を用いた辞典知識情報の抽出と組織化”, 電子情報通信学会論文誌 D-II, Vol. J85-D-II, No. 2, pp.300-307 (2002).
- 10) 土橋惇一, 荒木健治, “Web 文書を対象とした用語説明文抽出手法における抽出範囲の特定”, 情報処理学会研究報告, Vol.2006-NL-171, pp.37-42 (2006).
- 11) The WebSocket API, 入手先
<<http://www.w3.org/TR/2011/WD-websockets-20110929/>>
(2014.05.08)
- 12) Yahoo! JAPAN が提供するニュース WebAPI, 入手先
< <http://developer.yahoo.co.jp/webapi/news/> >
(2014.05.08)
- 13) Boilerpipe, 入手先
<<http://boilerpipe-web.appspot.com/>>
(2014.05.08)
- 14) Christian Kohlschütter, Peter Fankhauser, Wolfgang Nejdl, “Boilerplate Detection using Shallow Text Features,” Proceeding WSDM '10 Proceedings of the third ACM international conference on Web search and data mining, pp. 441-450, USA, 2010
- 15) キーフレーズ抽出 API, 入手先
<<http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html>>
(2014.05.08)
- 16) DeviceOrientation Event Specification, 入手先
<<http://dev.w3.org/geo/api/spec-source-orientation.html>>
(2014.05.08)
- 17) Web Speech API Specification, 入手先
<<https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>>
(2014.05.08)
- 18) Battery Status API, 入手先
<<http://www.w3.org/TR/battery-status/>>
(2014.05.08)