

# 未来へつなぐ新聞ビッグデータ

## Newspaper Big Data for the Future

服部俊紀 Shunki Hattori

### 1. 背景

新聞には数あるメディアの中でも話題の多様性や信頼性、一覧性など多くの利点がある。それにもかかわらず近年は手軽に情報を取得できるメディアへ購読者が流れている。そこで本プロジェクトでは、北海道新聞社の協力を得て過去 33 年の新聞記事データに情報技術を活用し、新聞記事との新しい出会いの場を作り出すことを目指した。本年度は 2 つのグループに分かれ、様々な馴染みのある形式で新聞記事が読めるニュースリーダーの開発、新聞をもとにした言葉遊びゲームの開発をテーマに活動した。

### 2. 課題解決のプロセスとその結果

#### 2.1 グループ A

##### 2.1.1 テーマに関する調査

グループ A では新聞記事が衰退した理由に、SNS やネットメディアが台頭し[1]、読者が移ったことをあげた[2]。また、SNS やネットメディアが新聞よりも優れている点は読みやすさではないかと考えた。新聞各社の動向を調査したところ、SNS やネット戦略に力を入れているものの、成功事例がニューヨークタイムズ[3]や日本経済新聞[4]などのごく少数であることがわかった。

##### 2.1.2 課題設定

そこでグループ A では、新聞各社が未だに取り組

んだことがない手法を用いて新聞記事にアクセスする人を増やすことを到達目標とした。具体的には、以下の 2 つの仮説を立て、これらを実現することで目標達成を目指した。

1. 新聞を馴染みのあるユーザインタフェース (UI) で読むことが出来れば、普段読まないユーザも興味を持つのではないか。
2. 新聞記事の形式を様々な日常の場面に適するように変更することで、新聞記事にアクセスする機会が増えるのではないか。

1 を実現するために多くの人に馴染みのあるサービスの UI を模すことにした。また、2 を実現するために 1 つの記事を長文、短文、長尺動画、短尺動画の 4 つの形式に変換することにした。これらを達成するための具体的な課題として、専用の UI 制作と記事データの加工、そのデータを保存するデータベースの構築があげられた。そこで、記事データの加工を担当する記事メディア処理班と、データベースと UI の開発を担当するアプリ開発班に分かれて活動した。

##### 2.1.3 記事メディア処理班の活動

記事メディア処理班では記事データの内容を変更することなく要約、動画化することを目標に活動を行った。まず記事データの前処理を行い、抽出型の要約アルゴリズムである LexRank[5]を用いて記事データの自動要約プログラムを制作した。その後、各記事に適した画像を探し出し、要約後の読み上げ

データと字幕に結合させるプログラムを制作した。  
これにより、要約文や動画の制作過程の一部を自動で行うことが可能になった。

### 2.1.4 アプリ開発班の活動

アプリ開発班では、まず提供データの表示に適するネットメディアを議論し、道新電子版、Twitter、YouTube、TikTok の4つを参考にして制作することにした。その後、各形式を記事一覧ページと個別ページの2層に分け、実装を行った。アプリ全体のシステムは図1の構成を取った。UI開発ではフレームワークにFlutterを採用し、状態管理にはProviderというライブラリを使用した。データベース構築にはFirebaseを用いて、加工された記事データを形式ごとに保存・取得できるように専用の属性やIDを付与した。

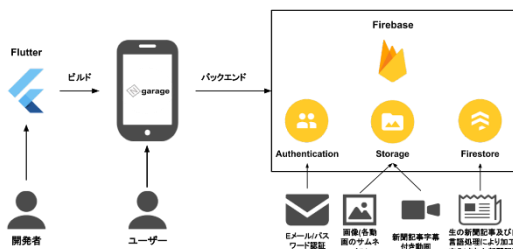


図1. N garage アーキテクチャ図

### 2.1.5 最終成果物

最終的な成果物として、慣れ親しんだUIで新聞記事を読むことができるスマホアプリ「N garage」を制作した（図2）。このアプリは、Original、Short Text、Video、Short Video という4つの形式を持ち、タップやスクロールをするだけで表示形式の変更ができる。これにより、ユーザが好みの形式で情報を取得することや異なる形式で読み比べることを可能にした。

大学生10人を対象とした本アプリの評価実験では、アプリの機能面や記事メディア処理に課題が見つかったものの、馴染みのある形式による表示や日常の様々な場面で利用できる可能性を高く評価された。

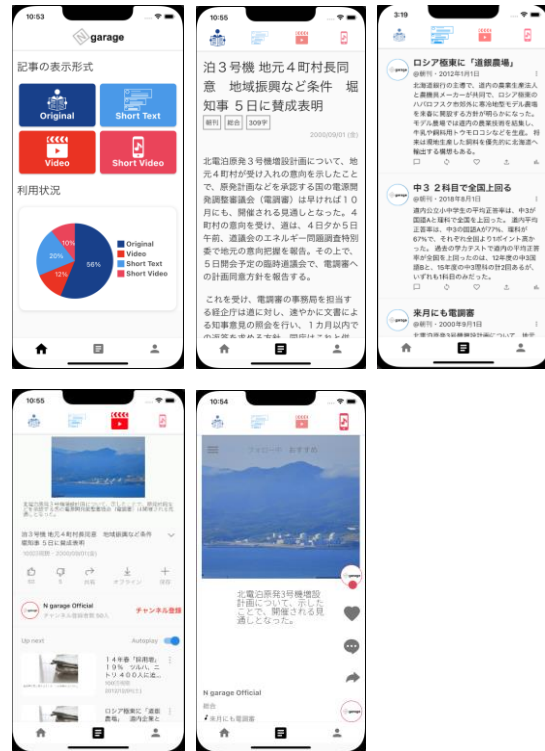


図2. ホームページ（左上）、Original ページ（中央上）、Short Text ページ（右上）、Video ページ（左下）、Short Video ページ（中央下）

## 2.2 グループB

### 2.2.1 テーマに関する調査

まず、グループBでは新聞には信頼性があると言われているにもかかわらず購読者数は減少していることを調査した[6][7]。そして、新聞社が行っているデジタル化などの新規事業について調査した[8]。また、ゲームの効果として付随学習があることについても調査した[9]。

### 2.2.2 課題設定

前述した背景からグループBでは、新聞記事との新しい出会いの場を創り出すことを目的として言葉遊びのスマホゲームを制作した。言葉データの素材として新聞記事から年度ごとの特徴語を抽出して利用することで、言葉遊びを通して過去の北海道の時事ニュースやトレンドワードの背景に触れる機会を創り出すことを試みた。そこで、制作にあたり辞書構築班とゲーム制作班に分かれて活動を行った。

### 2.2.3 辞書構築班の活動

辞書構築班では、まず新聞記事データをゲームに使用する形にするための分析を行った。はじめに CSV 形式の新聞記事データに含まれる<>や\*といった、実際の新聞には掲載されていないが、北海道新聞社内で使用されている文字列を正規表現で除去した。形態素解析エンジンには MeCab[10]を使用した。通常の MeCab 内の辞書では「新型コロナウイルス」などといった固有名詞に対応しておらず、抽出できなかったため、mecab-ipadic-NEologd [11]と連携させて辞書の性能向上を図った。その結果、1つ1つの名詞の出現回数を各年代の月ごとに計測することができた。しかし、この方法だけでは特定の年に出現回数が多くなる名詞を抽出できないため、ある年の出現回数が最大出現回数の半分以上を超えた場合に条件を満たしていると判定した(図3)。そして、その条件によって抽出した名詞をゲーム素材として使用した。

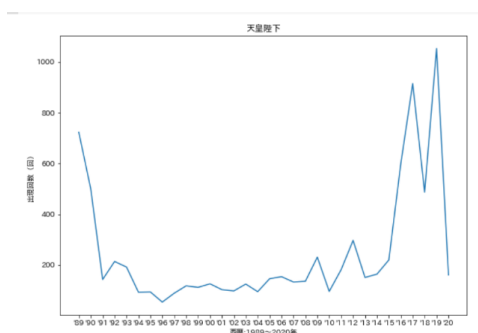


図3. 出現回数の最大値と最小値を比較するグラフ

### 2.2.4 ゲーム制作班の活動

ゲーム制作班では、はじめに全員で Unity での開発方法を学んだ。具体的には Unity の基本操作と Unity で用いているプログラミング言語である C# について学習した。ゲームを制作する際に作業を効率化するため、ゲームの機能を担当するグループと、ゲームのグラフィックデザインを担当するグループに分かれた。ゲームのグラフィックデザインを担当するグループでは画像編集ソフトの Adobe Photoshop/ Adobe Illustrator と 3DCG 作成ソフト

の Blender を用いた。ゲームの機能を担当するグループでは辞書構築班が作成したデータを用いてゲームを制作した。

### 2.2.5 最終成果物

最終成果物として、「NEW SUPER NEWSPAPER」という5つのゲームを統合させた言葉遊びゲームを制作した。各ゲームは「Hit And Blow Text」「8 パズル」「もじもじまぜまぜ」「ワードトレンドクイズ」「SWAPPING LETTERS」という名前がついており、ゲーム性もそれぞれで異なっている(図4)。それぞれのゲームでは、新聞記事から抽出した特徴のある名詞で構成された辞書を使用しており、新聞に記載されている単語を用いた問題を解くことができるほか、その単語に関する新聞記事を読むことができる。

ゲームの有用性を評価するため大学生10人を実験参加者とした実験を行なった結果、新聞記事がゲームに活用できていると回答したのは10人中8人だった。そのため、言葉遊びを通して過去の北海道の時事ニュースやトレンドワードの背景に触れる機会を創り出すことができたと考えられる。

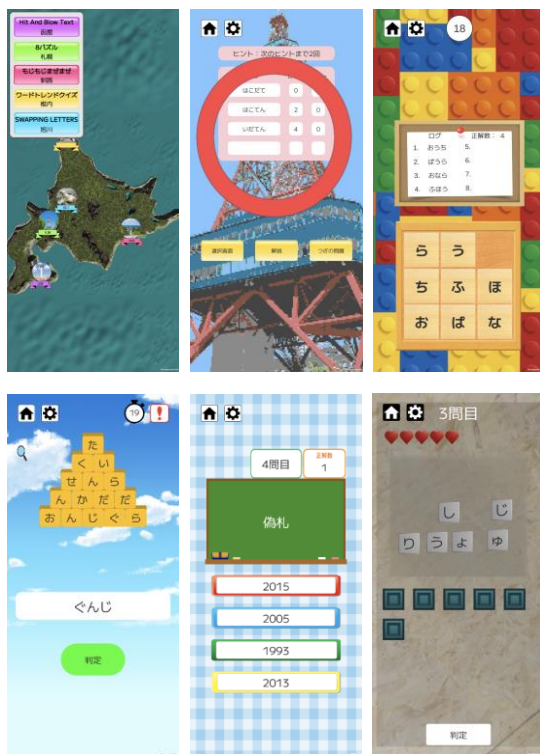


図4. ゲーム選択画面(左上)、Hit And Blow Text

(中央上), 8 パズル (右上), もじもじまぜまぜ (左下), ワードトレンドクイズ (中央下), SWAPPING LETTERS (右下)

### 3. まとめと今後の課題

本プロジェクトでは、新聞記事との新しい出会いの場を作り出すことを目的に2つのグループに分かれて活動を行った。

グループ A では、慣れ親しんだ UI で新聞記事を読むことができるスマホアプリを開発した。このアプリは、1つの記事を4つの形式（長文、短文、長尺動画、短尺動画）で読めることから、従来の新聞紙や SNS では叶わなかった、ユーザの特性や今いる環境に応じて記事をシームレスに変更することを可能にした。また、各形式を既存の代表的なメディアを参考に制作することで、ユーザが直感的に操作することも可能にした。評価実験の結果、動画の完成度向上や検索機能の実装などが課題として残ったが、コンセプトや実装できた機能は高く評価された。

グループ B では、「NEW SUPER NEWSPAPER」という5つのゲームを統合させた言葉遊びゲームを制作した。このゲームは、新聞記事から抽出した特徴のある名詞で構成された辞書を使用しており、新聞に記載されている単語を用いた問題を解くことができるほか、その単語に関する新聞記事を読むことができる。このゲームの有用性に関する評価実験を行なった結果、特徴的な単語の抽出方法やゲームのロード時間の改善などに課題があるが、言葉遊びを通して過去の北海道の時事ニュースやトレンドワードの背景に触れる機会が創られたことを確認することができた。

### 謝辞

本プロジェクトでは北海道新聞の過去33年分の新聞記事データを活用した。プロジェクトの構想やシステム試作において北海道新聞函館支社長の三浦辰治氏に多大なるご協力を頂戴したので感謝申し上げます。

### 参考文献

- [1] 経済産業省, 特定サービス産業動態統計調査, <https://www.meti.go.jp/statistics/tyo/tokusabido/index.html> (2021. 07. 11 取得)
- [2] 公益財団法人新聞通信調査会, 第13回メディアに関する全国世論調査, <https://www.chosakai.gr.jp/project/notification/> (2021. 07. 10 取得)
- [3] なぜNYタイムズは新聞のデジタル化に唯一成功した? 紙の売上がまもなく抜かれる, Money Voice, <https://www.mag2.com/p/money/928005> (2021. 07. 10 取得)
- [4] 江村亮一, 日経電子版, これまでの歩みと今後の成長へ向けた課題, [https://www.jstage.jst.go.jp/article/jkg/68/9/68\\_446/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jkg/68/9/68_446/_pdf/-char/ja) (2021. 11. 17 取得)
- [5] G. Erkan and D. R. Radev, LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research, 22, pp.457-479, 2004.
- [6] 総務省, 令和元年版 情報通信白書, pp.98-100, 2018.
- [7] 日本新聞協会, Pressnet 一般社団法人 日本新聞協会 新聞の発行部数と世帯数の推移, <https://www.pressnet.or.jp/data/circulation/circulation01.php> (2021. 7. 10 取得).
- [8] 松本恭幸, 地方紙の現状と課題, 情報の科学と技術, 68(9), pp.440-445, 2018.
- [9] 藤本徹, ゲーム学習の新たな展開, 放送メディア研究, No.12, pp.242-243, 2015.
- [10] Taku Kudo, MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/> (2022. 01. 12 取得)
- [11] NEologd, Mecab-ipadic-neologd, <https://github.com/neologd/mecab-ipadic-neologd> (2022. 01. 12 取得)