

新聞ビッグデータ

知的好奇心を刺激

ビッグデータを活用

本プロジェクトでは、「新聞ビッグデータから「何か」を生み出す」というテーマに基づき、知的好奇心を刺激する「ViewPicks」というアプリケーションを開発しました。開発に際し、北海道新聞社より提供を受けた2008年7月1日から2020年12月31日までの新聞テキストビッグデータを自然言語処理で解析し可視化することにより時代の動向を語彙という観点からアプ

新聞離れが深刻化

新聞の発行部数は日を追うごとに減少傾向にあります。地方に着目した情報が減るといことは、地元の政治の情報などの地域に密着した情報を得る機会が減少するということにも繋がります。新聞では政治、経済、国際情勢などの幅広い分野の情報が掲載されています。

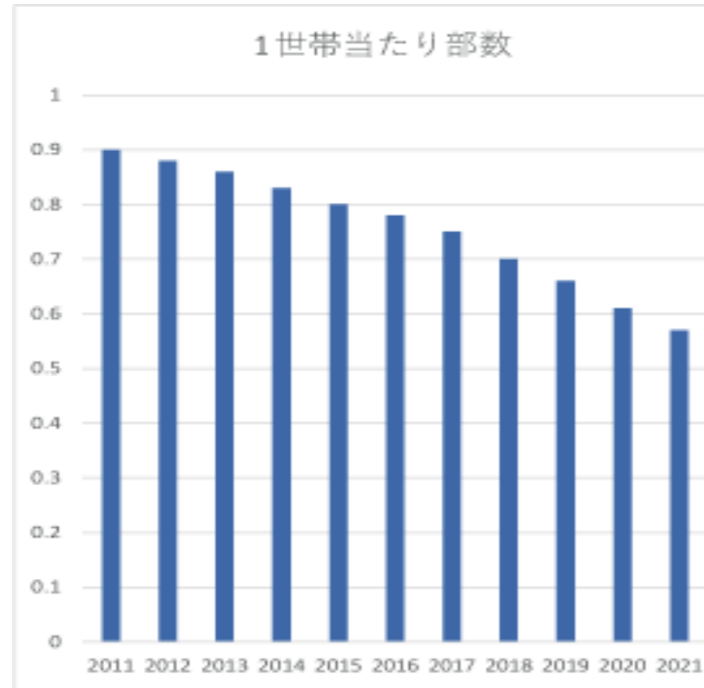


図1: 1世代あたりの発行部数

一方、SNSではユーザーが見たい情報だけを取り入れることができ、新聞とは様態が異なります。近年のメディアの特性、特にTikTokやTwitterなどのSNSにおいてユーザーの興味に合わせた最適化された内容が表示される傾向にあります。

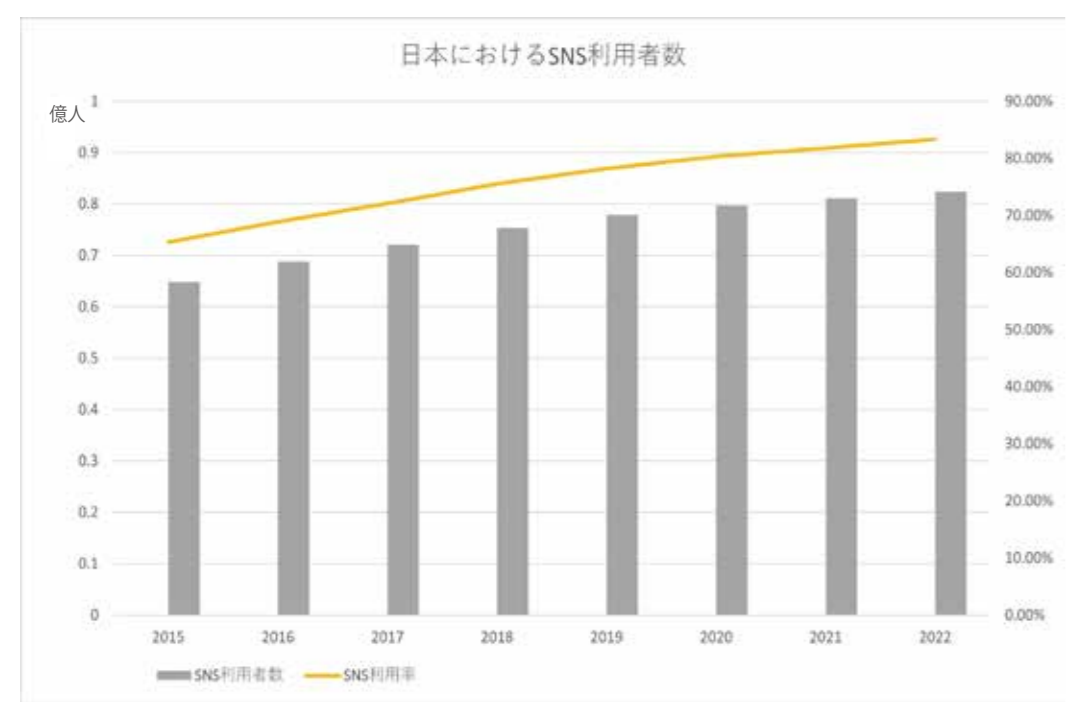


図2: 日本における SNS 利用者数

かし、新聞では政治、経済、国際情勢など幅広い情報が掲載されています。また、新聞の情報は一次情報であり、SNSのように情報が拡散され、デマ情報を受け取る危険性が少ないです。このような観点から、新聞のデータを可視化した過去の情報を観察すると共に新聞に眠る情報の面白味を感じていただけます。

ビッグデータを可視化し未来へ残す

多様な時事に富んだデータをアプリで閲覧

公立はこだて未来大学の学生8名がアプリを開発

「ViewPicks」というアプリケーションでは、新聞のテキストデータから新聞内における情報の強弱や関係性をインタラクティブに可視化し知的好奇心を刺激する体験を提供します。また「眺めてるだけ」ではなく「眺めてるだけ」のような考えを持っていただけような仕掛け作りを施しました。アプリケーションのプラットフォームにWebを選択し、JSON形式のデータに新聞テキストを加え、利用しています。JSONデータを基にword cloudやcircle packingなどの情報の可視化を実現しています。実装に当たってはJavaScriptのD3.jsというライブラリを用いて実装を行いました。



写真1: アプリケーションのホーム画面

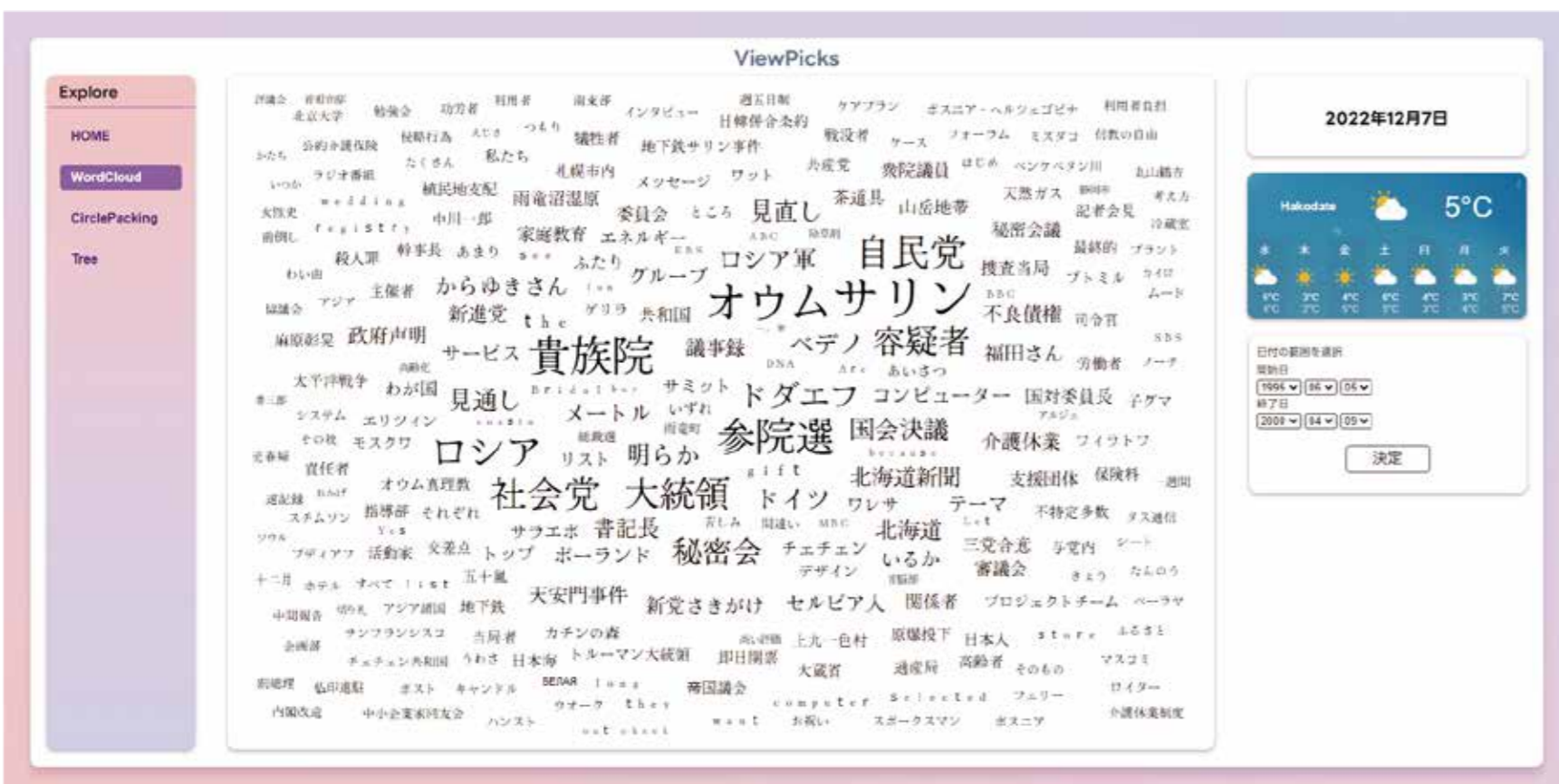


写真2: アプリケーションの可視化画面 (Word Cloud)

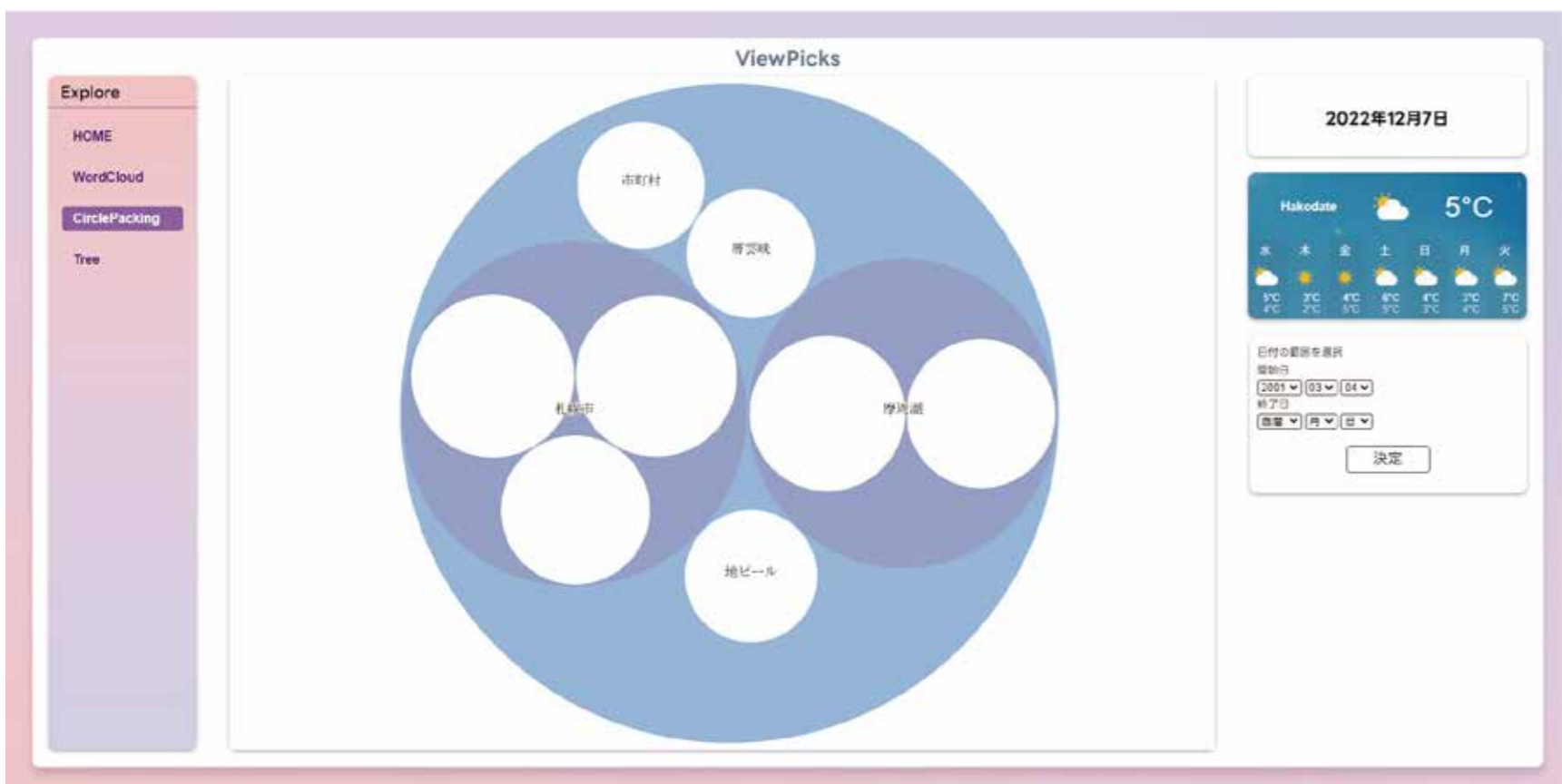


写真3: アプリケーションの可視化画面 (Circle Packing)

Word Cloud

指定された日付の区間での単語の頻出度を表現

- ・過去の話題を知ることが可能
- ・ジャンル関係なしに表示
- ・大きさで直感的に理解可能

今後の展望

- ・ジャンルによる色分け
- ・単語をクリックすると関連した記事の表示

Circle Packing

指定された日付の区間での単語間の関連度を入れ子構造で表現

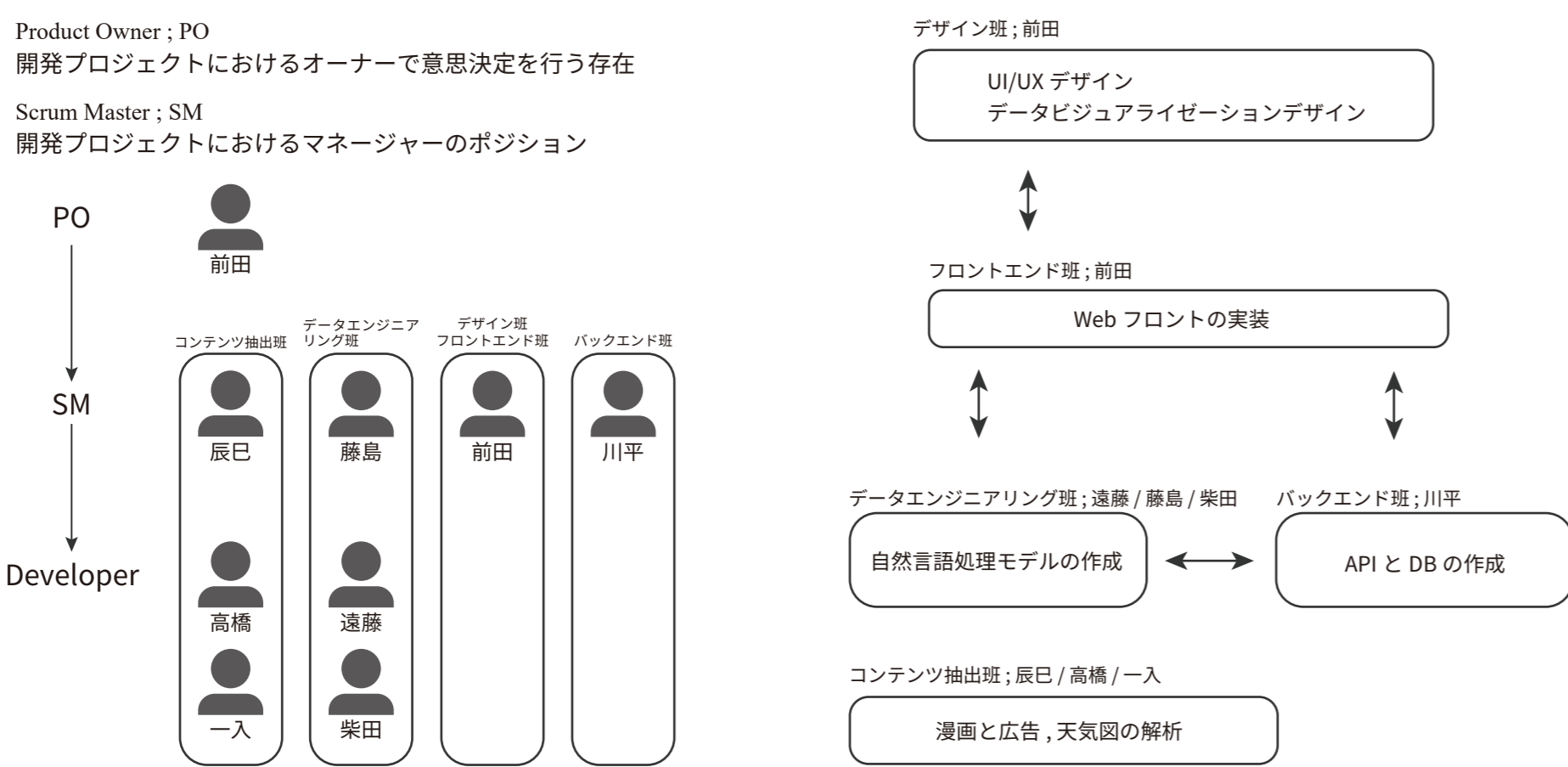
- ・円のサイズは関連度の量的データを示す
- ・話題の渦中を知ることができる

今後の展望

- ・単語から記事の関連性を推定し表示

開発の効率化に成功

Webアプリケーションの開発に際しスクラム開発を実践しました。デザイン班、フロントエンド班、バックエンド班、コンテツツ抽出班、データエンジニアリング班、自然言語処理班、APIとDBの作成班、漫画と広告、天気図の解析班の体制で開発を行いました。各班はDeveloperも兼任しました。各班は目的は短期間で開発において効率的に開発を行う必要があるため、後期から各班の役割を明確にすることで開発の進捗を把握し、約1ヶ月で制作する必要があった。各グループでタスクを分散し、お互いに認識をすり合わせながら進めた結果、短期間で効率的に開発を行うという目的を達成することができました。



活動計画

- ・新聞について学習
- ・アイデア検討
- ・自然言語処理や他の技術について学習
- ・技術を学んだ上で実現可能性のあるアイデアを検討
- ・プロトタイプの開発
- ・中間報告
- ・夏季休暇の開始
- ・書籍や論文、インターンを通して知識と技術の習得
- ・夏季休暇の終了
- ・アイデアの再検討
- ・GitHub ワークショップ
- ・各班に分かれプロトタイプを作成
- ・web アプリケーションの開発
- ・最終成果発表会
- ・報告書の提出
- ・プロジェクト学習課外発表会
- ・次年度へ