member : Akira Maeda Haruto Endou Satoshi Kawahira Kouki Shibata Naoya Tatsumi Yuuki Hitoshio Kairi Fujishima Teachers : Kengo Terasawa Rui Sakaida Yasuyuki Sumi Noyuri Mima Adviser : Tatsuharu Miura (The Hokkaido Shimbun Press) Collaboration : The Hokkaido Shimbun Press

Newspaper Big Data for the Future

9 December 2022 Final Outcome Report

Using big data to stimulate intellectual curiosity

This project developed a web application called "ViewPicks", which visualises newspaper text data and stimulates intellectual curiosity, based on the theme of "creating something" from newspaper big data. The web application analyses and visualises newspaper text data from 1 July 1988 to 31 December 2020 using natural language processing, and displays the trends of the times in terms of vocabulary.

Newspapers have a number of advantages over other media, such as diversity of topics, reliability and listability. The aim is to provide the value of information that lies dormant in newspapers through a tool that enables users to interact with the language used in newspapers. The vision is to provide the meta-information that lies dormant in newspapers and to increase interest and involvement in newspapers.

Newspapers are becoming increasingly popular.

The circulation of newspapers is declining by the day. The decline in local information means fewer opportunities to obtain locally relevant information, such as information on local politics. Newspapers provide information on a wide range of topics, including politics, economics and international affairs.

There is less risk of receiving them. From this perspective, visualising newspaper data and observing past information stimulates intellectual curiosity and gives an interesting insight into the information lying in newspapers.

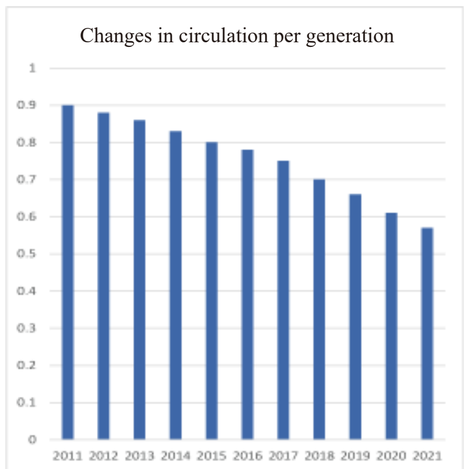


Figure 1 : Changes in circulation per generation

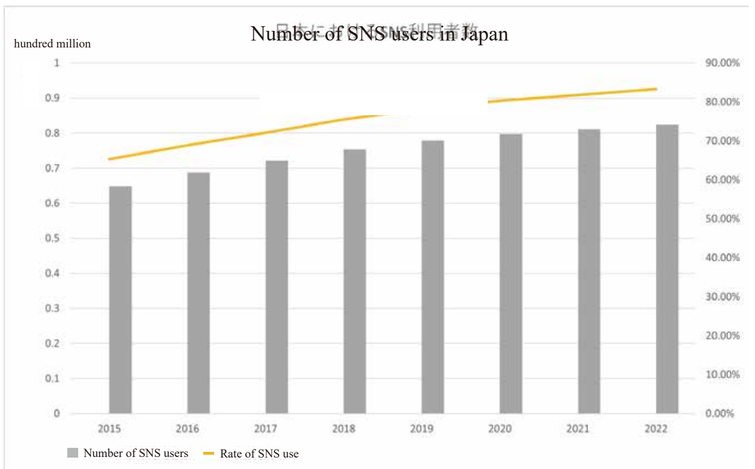
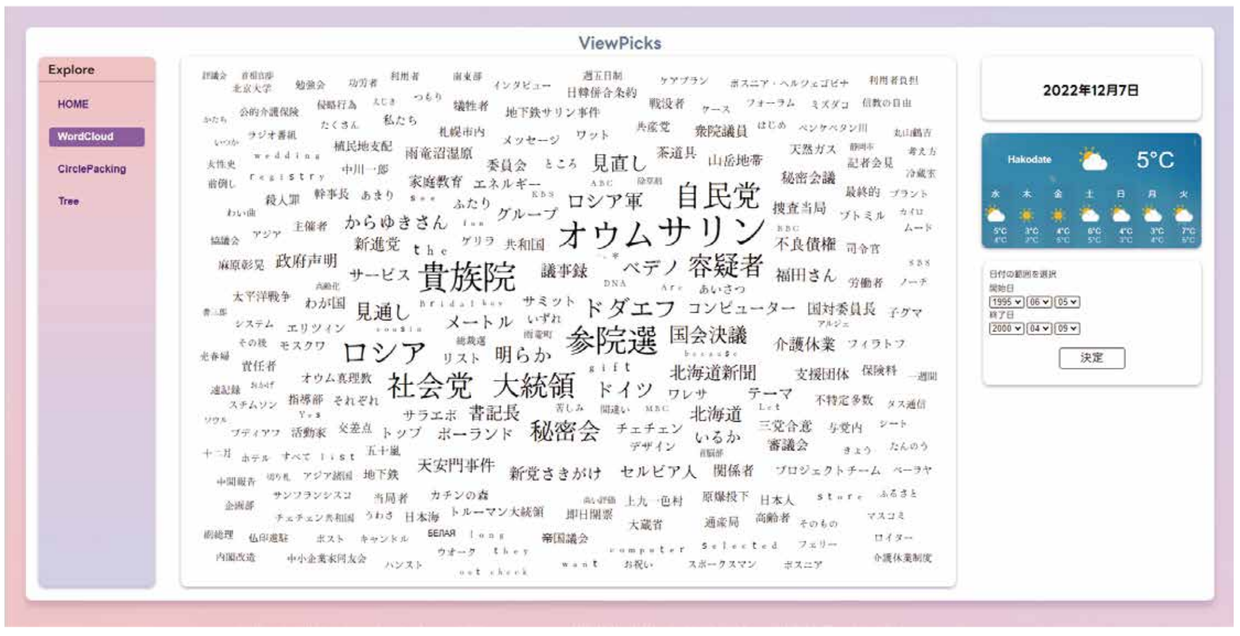


Figure 2 : Number of SNS users in Japan

View a diverse range of current data on the app

The ViewPicks application provides an experience that stimulates intellectual curiosity by interactively visualising the strengths, weaknesses and relationships of information within a newspaper based on the newspaper's text data. It is also designed to make people think that 'just looking at it is a bit interesting'. The application platform is the web, and newspaper text is processed into json format data, which is used to visualise information such as word clouds and circle packing based on json data. The implementation was carried out using the JavaScript library D3.js.

The design team was responsible for the design in UI, UX and data visualisation. The team designed the loading screen in consideration of the processing time when data is passed from the back-end, and realised the design of the application with an awareness of OOUI through the approach of animation. In order to realise an interactive system, we used the concept of components and arranged information in chunks, aiming for an easy-to-understand information design while ensuring an interactive look and feel. We implemented the web application using HTML, CSS and JavaScript, and utilised a JavaScript visualisation library called D3.js to create an interactive system. The development was carried out while performing task management using GitHub. By managing tasks and collaborating with other groups, we were able to efficiently develop an interactive system in a short period of time.data. The implementation was carried out using the JavaScript library D3.js.



Word Cloud

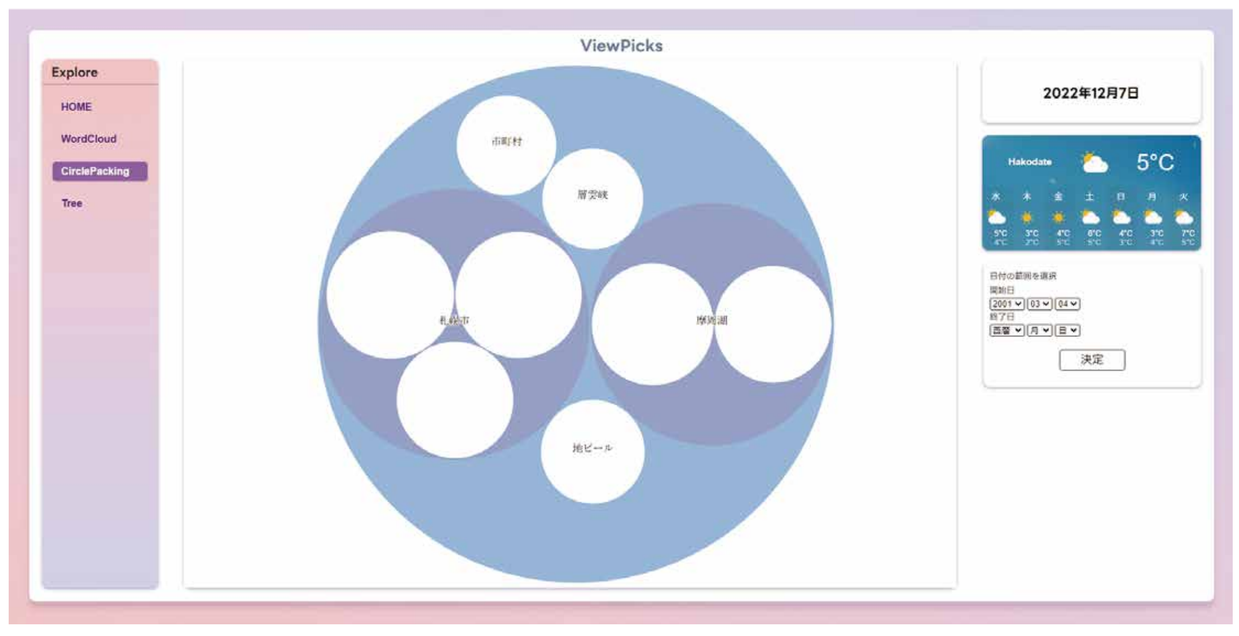
Expresses the frequency of a word in a given date interval

- Allows to know the topic in the past.
- Displays regardless of genre
- Intuitive understanding of the word by its size

Future prospects

Colour-coding according to genre

Clicking on a word displays related articles.



Circle Packing

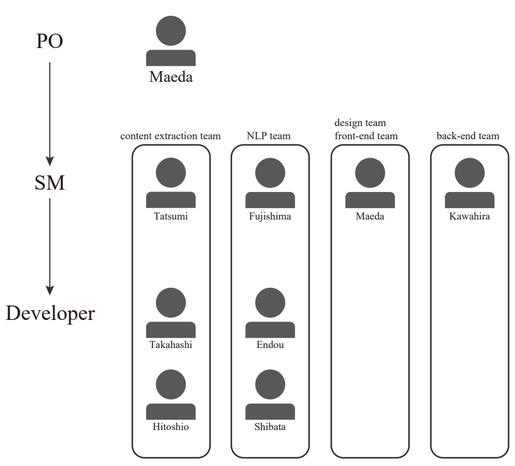
Nested structure representing the degree of association between words in a given date interval.

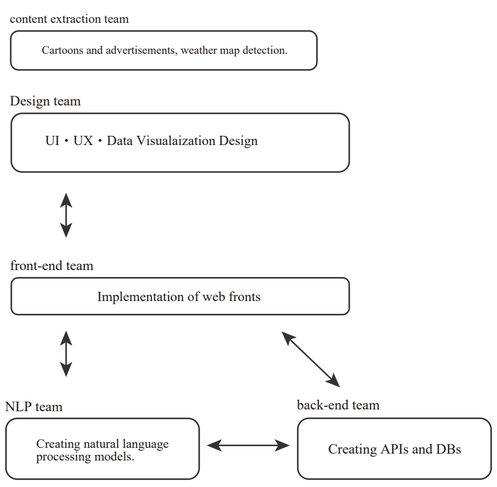
- The size of the circle indicates the quantitative data of the relevance.
- The size of the circle indicates the quantitative data of the relevance.

Future perspectives

The relevance of articles can be estimated from words and displayed.

Successful development efficiency





Scrum development was practised in the development of a web application. We divided the task into five teams - the design team, front-end team, back-end team, NLP team and content extraction tean - and proceeded with development while distributing the tasks and reconciling information through activity reports each time.

As we started developing the web application in the second semester, we needed to produce it in about two months. As a result of distributing the tasks among the groups and ensuring mutual recognition, we were able to achieve the objective of efficient development in a short period of time.