Образ Сахалина в японских СМИ

Студенты: Якушева Елена, Цыбикова Екатерина, Лаптева Ксения, Ников Максим, Шестопалов Тимофей

Тьюторы: Ушак Ксения, Вещикова Оксана

Исследовательский вопрос

Как сегодня представлен Сахалин в Интернет-пространстве Японии, в том числе в туристических блогах?

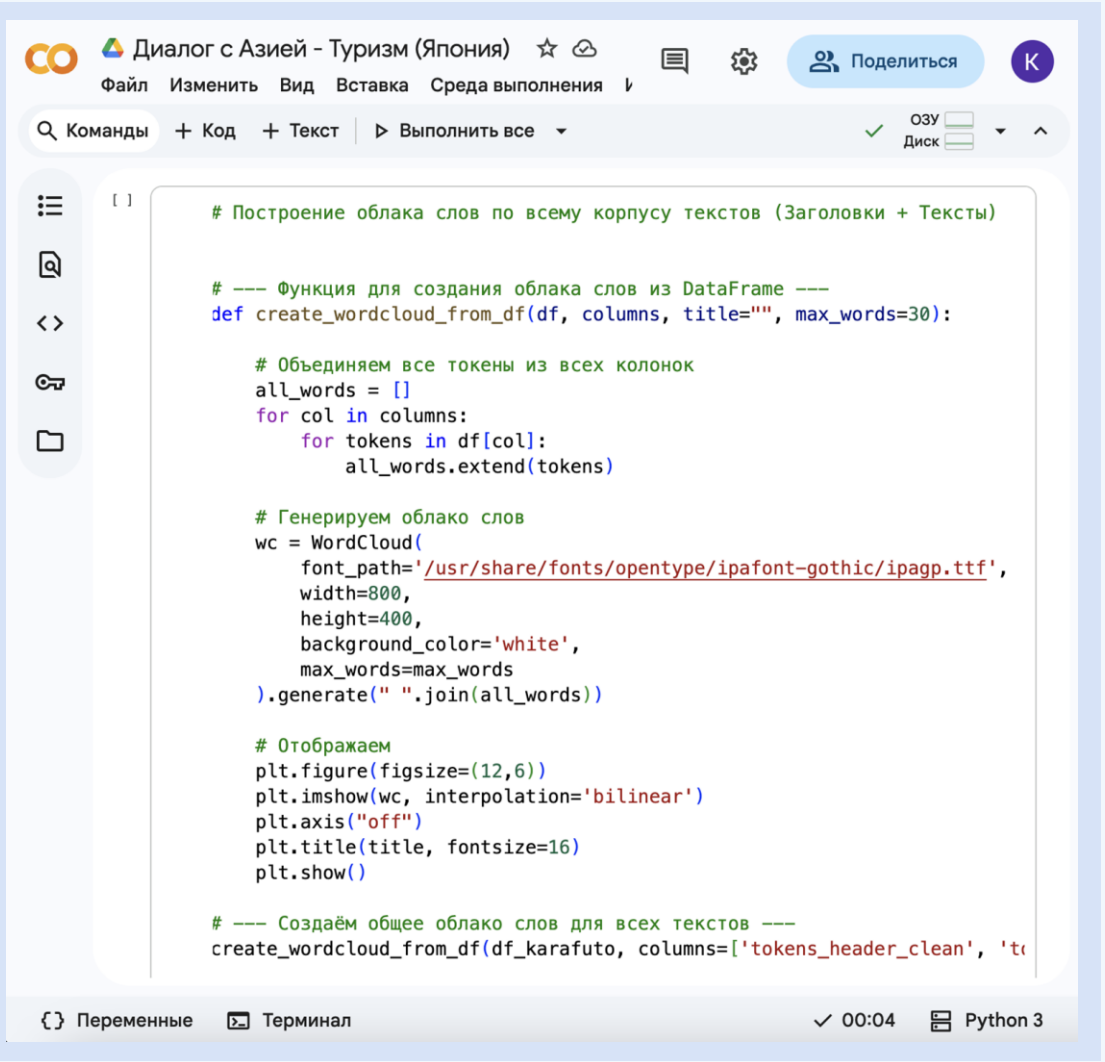
Актуальность

Туризм - приоритетное направление развития Сахалинской области

Источники

Более 500 статей

- news.yahoo.co.jp
- 4travel.jp

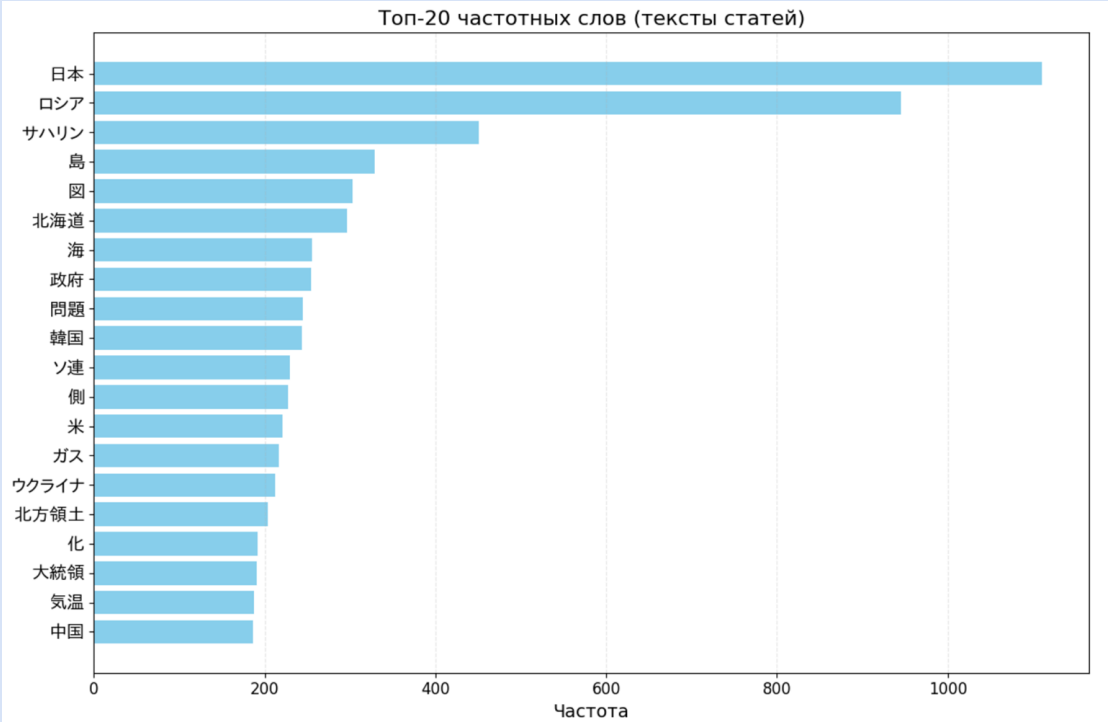


Методы исследования данных

Парсинг	Автоматизированный сбор данных с веб-сайтов. Программа открывает страницы, анализирует содержимое и извлекает нужную информацию (заголовки, даты, тексты, ссылки), экономя время на ручном сборе данных.
Токенизация	Разбиение непрерывного текста на отдельные смысловые единицы (токены): слова, знаки препинания, числа.
Частотный анализ	Используется для выявления определения наиболее употребительных слов; ключевых тем и понятий.
Анализ совместной встречаемости (co-occurence analysis)	Метод помогает понять какие слова в тексте часто встречаются вместе, т.к. это явный признак того, что они связаны между собой.
LDA	Метод тематического моделирования, который помогает автоматически выявлять скрытые темы в большом наборе текстов.

Анализ по поиску “サハリン” (Сахалин)

Частотный анализа



LDA : Темы

- Международные отношения
- Проблема Северных территорий
- Энергетика и международная торговля
- Коренные народы
- Погода

Анализ по поиску “樺太” (Карафуто)

LDA : Темы

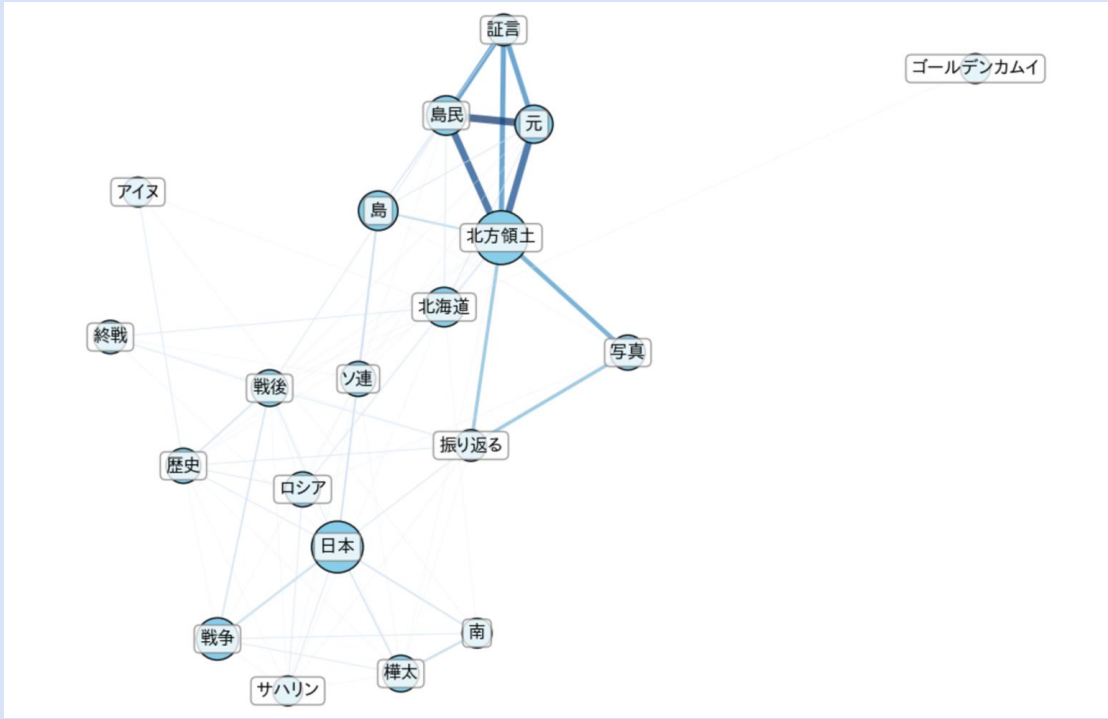
История Японии 20 века

Древние коренные народы

Манга и аниме

Погода в историческом контексте

Анализ совместной встречаемости



Анализ туристических заметок

Облако слов

Выводы

Главный туристический интерес – японское прошлое Сахалина

Целевая аудитория – японцы возрастом 60+

Сообщение не только паромом из Хоккайдо, но и в рамках круизов

Японцы активно пользуются электронной визой при въезде в Россию