Developing new library services using AI (machine learning) : an introduction to the *Next Digital Library*

Tahee ONUMA, Chief, R&D Office
National Diet Library, Japan
18 September 2021

CC BY

The R&D Office

- Founded in October 2011
- Conducts research into cutting-edge technologies in collaboration with researchers from outside the NDL
- Releases experimental or newly-developed services on ***the NDL Lab*** website(⇒ next slide), such as:
  - *The Next Digital Library* *today's main topic
  - *NDC Predictor*
  - *Japan Search*
  - *Bibliographic Information Retrieval and Visualization System*

  …

NDL Lab
National Diet Library, Japan

The NDL Lab

- Website for conducting field trials of new library services
- First launched in May 2013, renovated in March 2020
- Recent activities focus on machine learning

https://lab.ndl.go.jp/

NDL Lab
National Diet Library, Japan

Why machine learning?

- Machine learning helps improve
  - searchability based on content

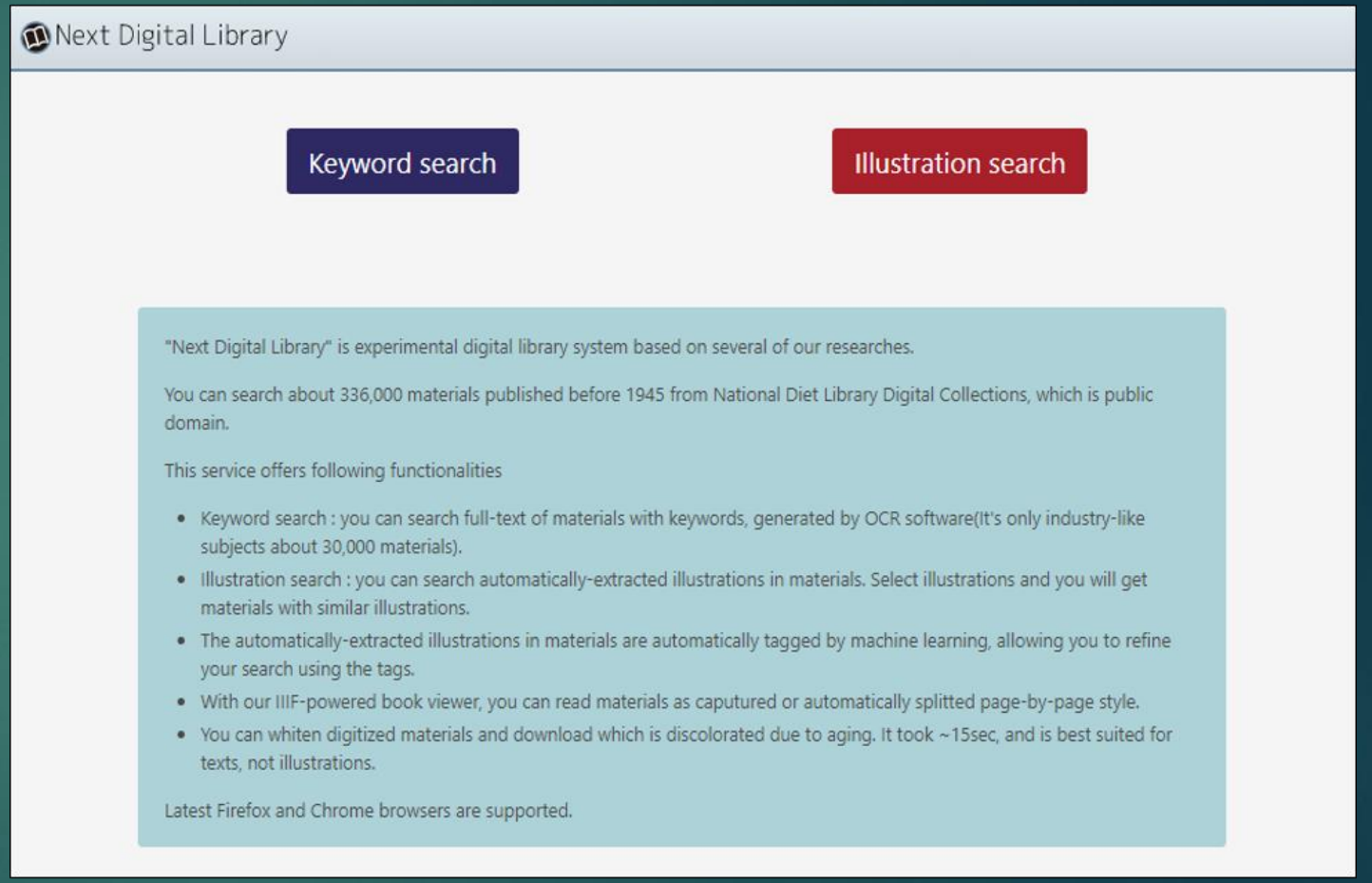
    Keyword full-text search and illustration search
  - readability of digitized images

    Bleaching, automatic image processing
- Functionality like this is implemented and tested on the *Next Digital Library*.

NDL Lab
National Diet Library, Japan

The Next Digital Library

- Launched in 2019 by the R&D Office
- A testing ground for experimental functionality developed by the NDL Lab
- Can be used to search the NDL Digital Collection for documents in the public domain
- Two search modes:
  - Keyword full-text search
  - Illustration search



https://lab.ndl.go.jp/dl/

NDL Lab
National Diet Library, Japan

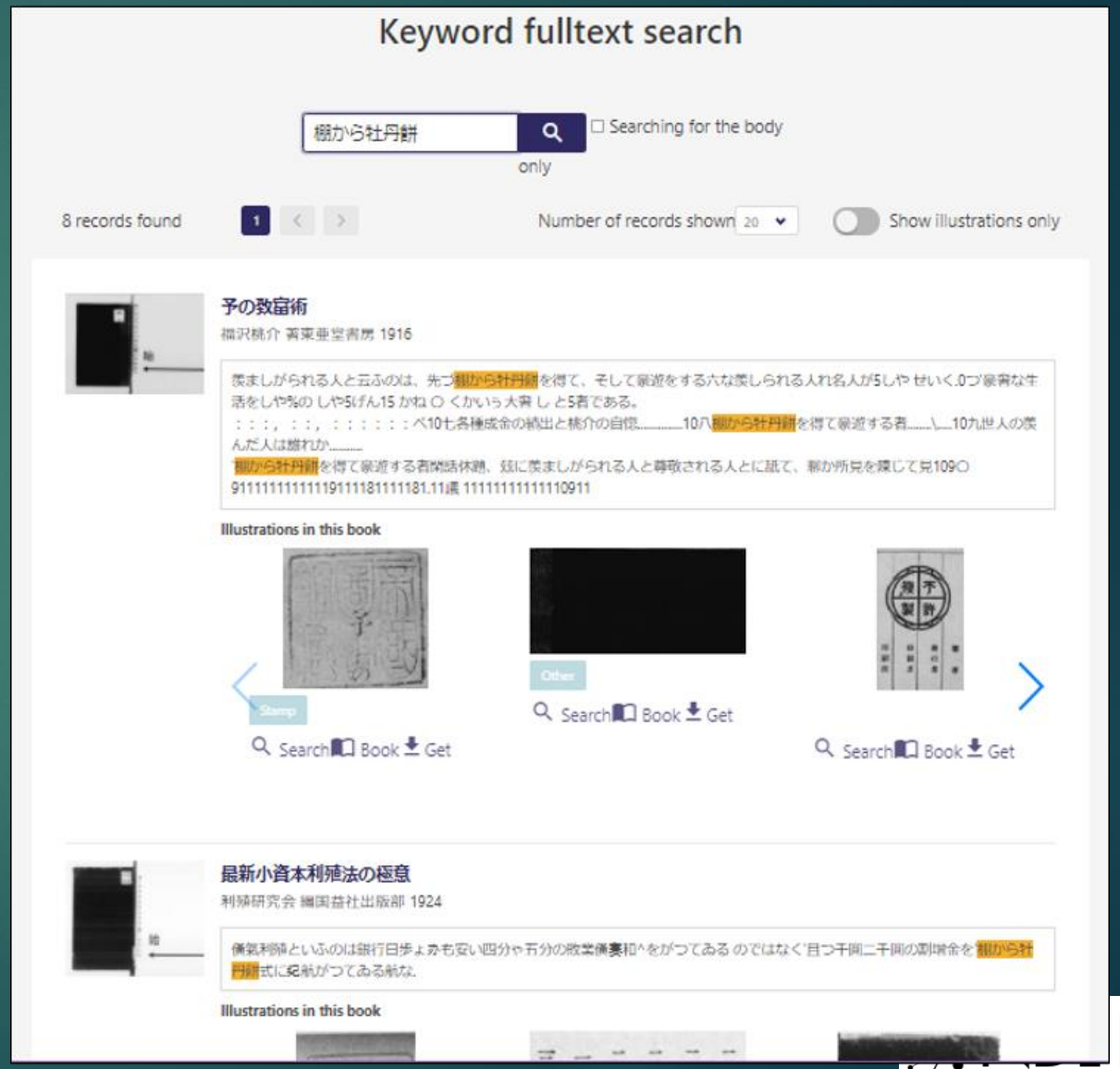
Keyword full-text search

- 30,000 items on industrial and commercial subject matter

  (Category 6 of the Nippon Decimal Classification)

- Searches full texts generated by OCR

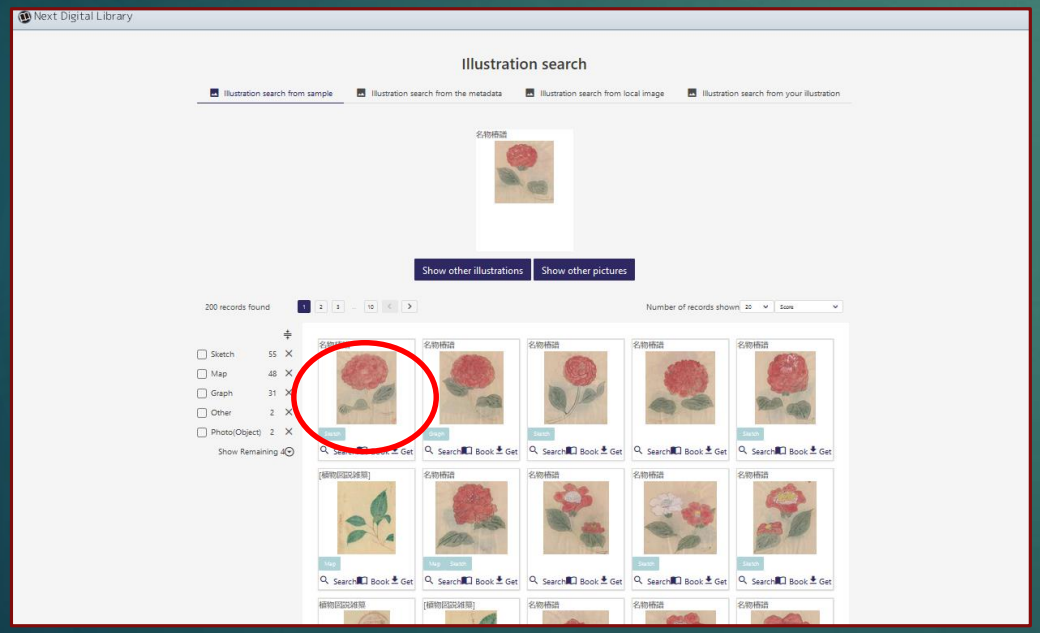
- Identify and access individual pages that include keywords

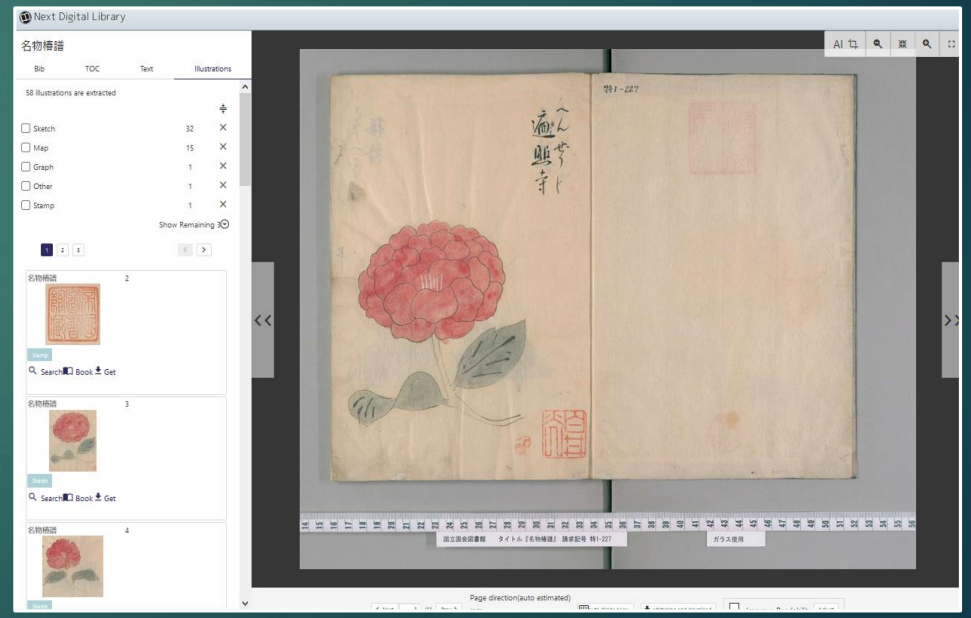


Lab
National Diet Library, Japan

Illustration search (1)

- Search for similar images
- 23,000,000 public-domain images from 336,000 digitized books, rare books, and historical materials in the *NDL Digital Collection*
- Increased potential of information exploration that goes beyond keyword search

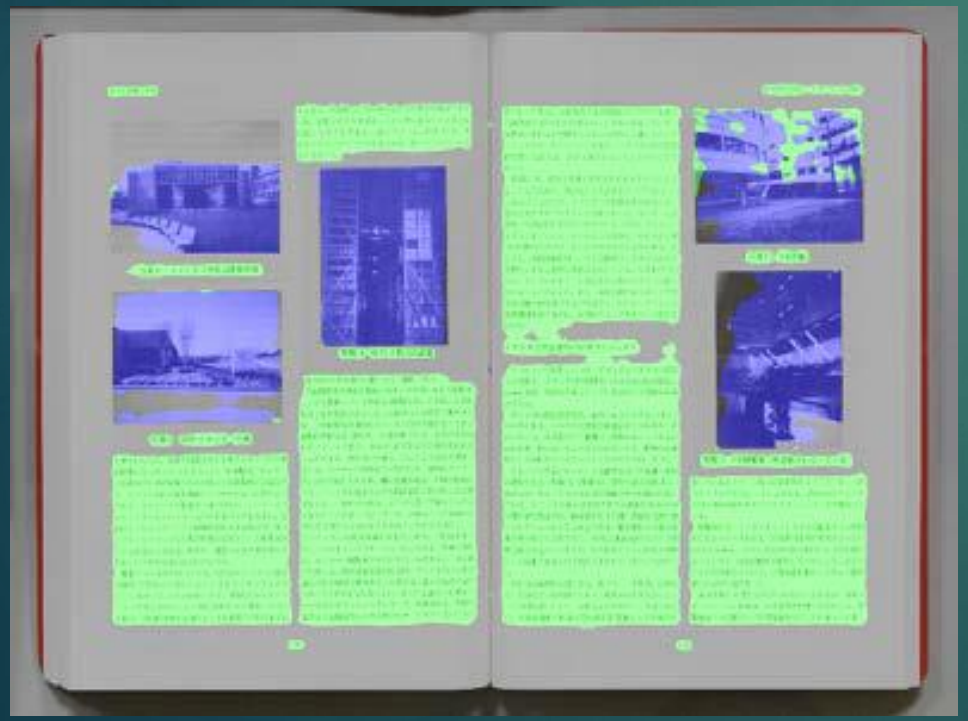


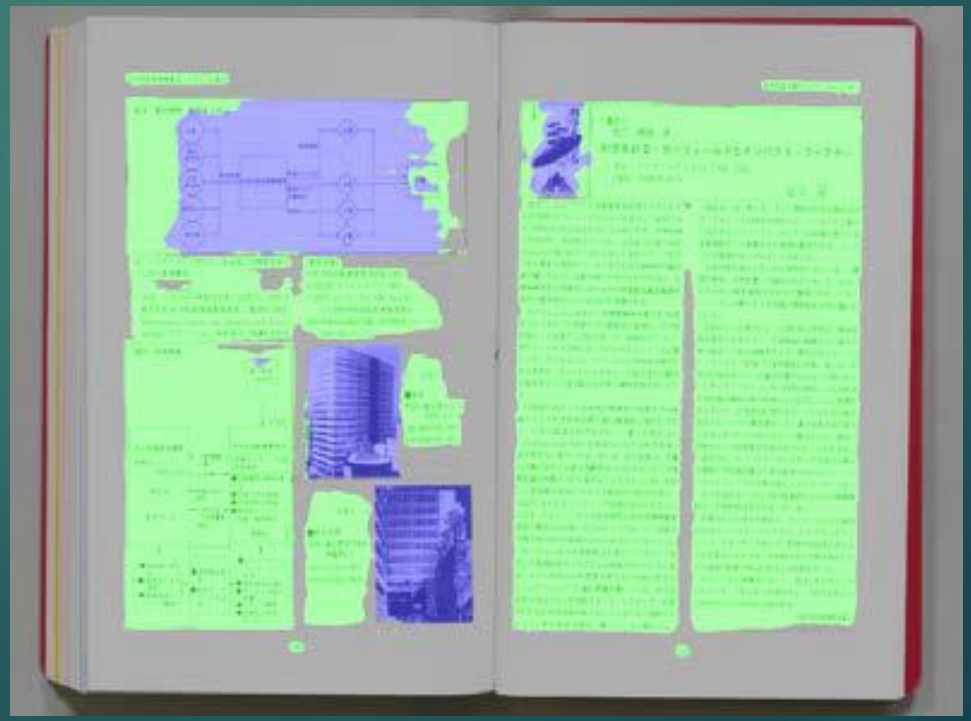


NDL Lab
National Diet Library, Japan

Illustration search (2)

- Image extraction: deep learning method “semantic segmentation“
  → Will change in the near future
- Feature extraction: trained using ImageNet dataset
- Search engine: vdaas/vald (https://github.com/vdaas/vald)



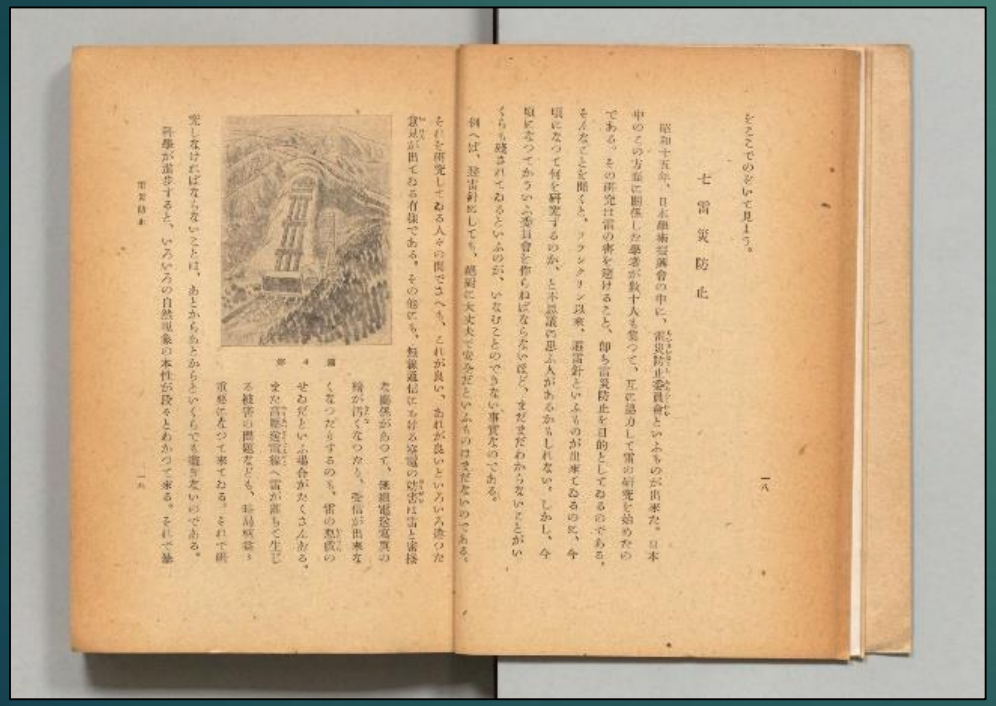


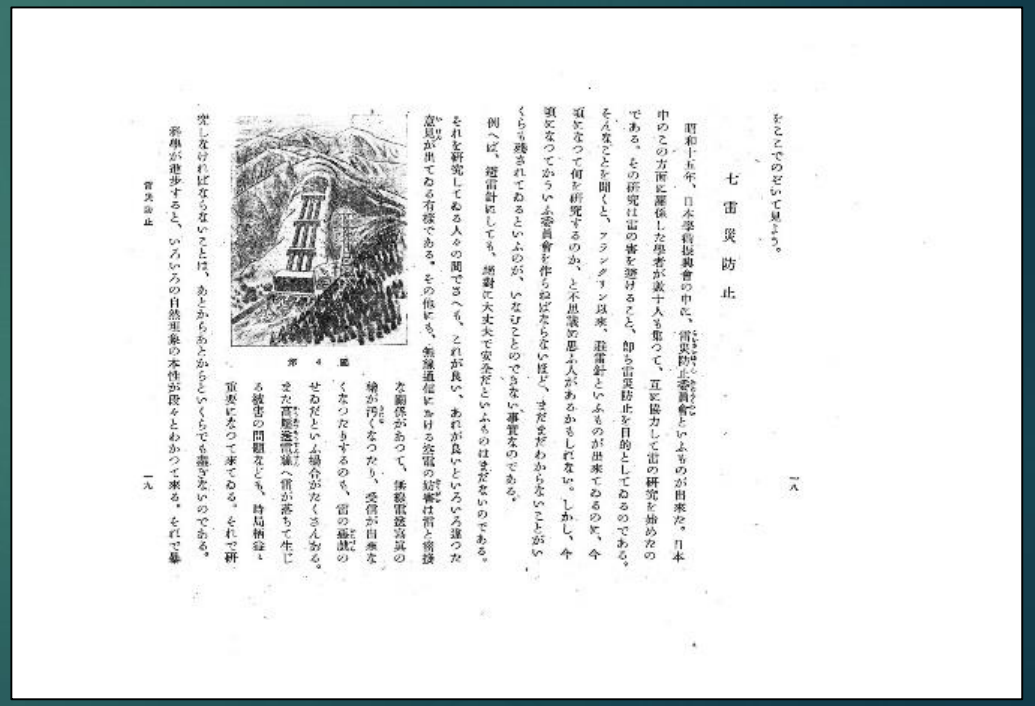
Blue regions indicate areas extracted to illustrations.

NDL Lab
National Diet Library, Japan

Background bleaching

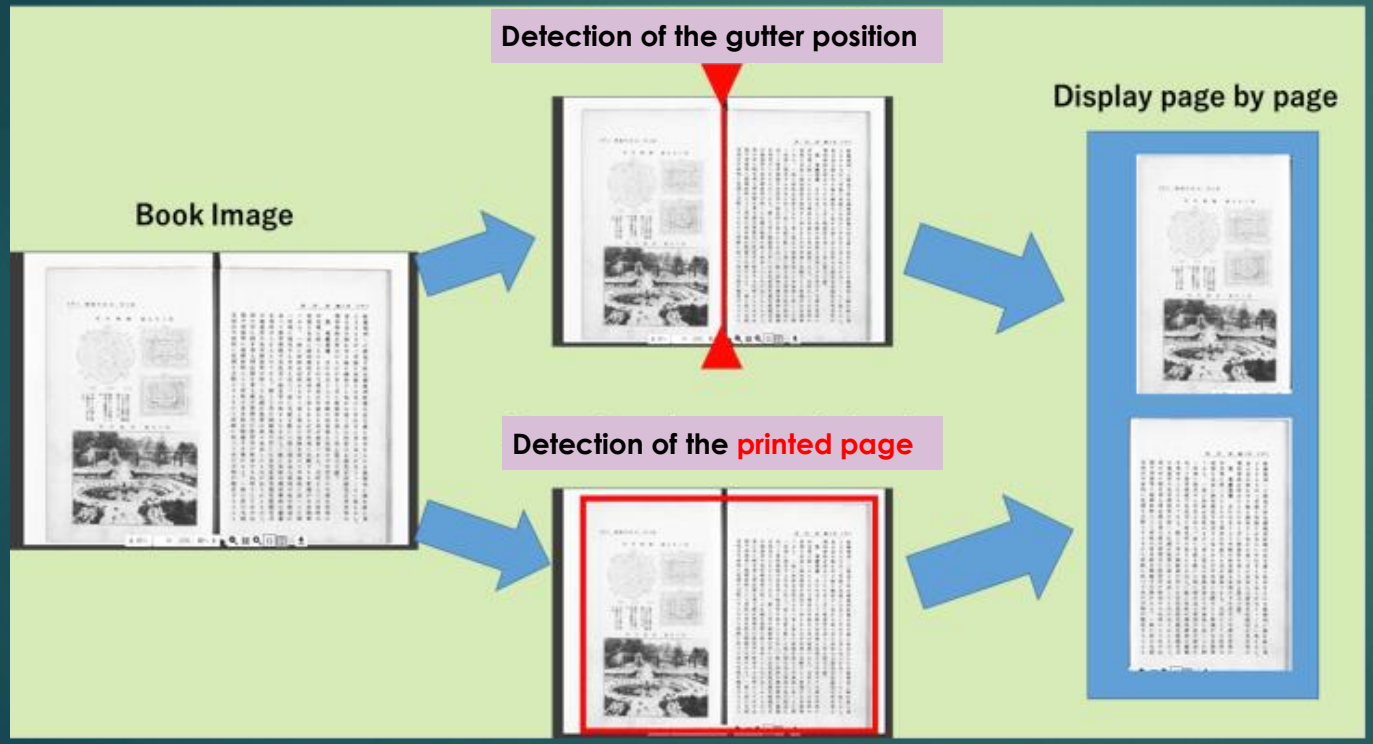
- Whitens pages discolored due to aging
- Improves readability
- Based on Neural Network model “pix2pix:GAN”





Automatic image splitting

- Automated splitting of pages and removal of margins
- Optimizes browsing on smartphones and tablets



NDL Lab
National Diet Library, Japan

Further application to other services

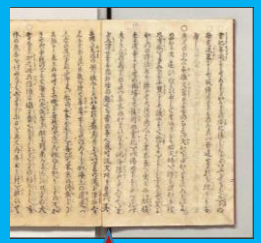
Research for the Next Digital Library

Detection of splitting position

A tool to assist quality inspections of digitized materials.

OK

Centered



NG

Forced out

Image search

Search thumbnail images in Japan Search (https://jpsearch.go.jp/)

JAPAN SEARCH BETA

Library services from the National Diet Library

NDL Lab
National Diet Library, Japan

Japan Search

- A national, integrated search platform for digital archives in Japan
- Launched August 2020
- Search 22.8 million metadata from 131 databases (August 2021)
- Similar Image Search function using technology tested on the Next Digital Library

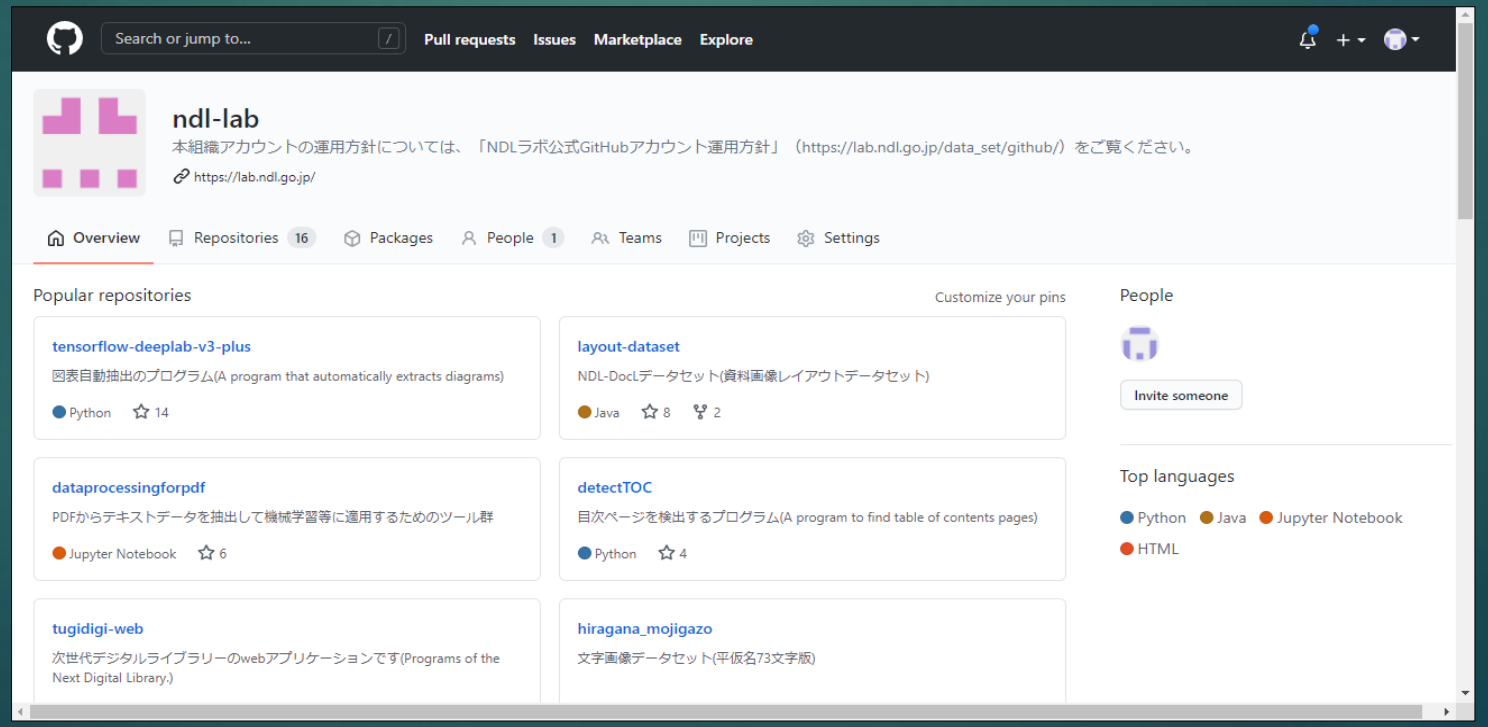
https://jpsearch.go.jp/

NDL Lab
National Diet Library, Japan

GitHub

- Provides access to programs (CC BY) and datasets (PD) created by the R&D Office
- A hub for the exchange of expertise with engineers from all over the world

https://github.com/ndl-lab



NDL Lab
National Diet Library, Japan

Ongoing projects

- Development of OCR software
- Preparing OCR-generated texts of digitized materials on the NDL Digital Collections
- Providing access to high-quality datasets

NDL Lab
National Diet Library, Japan