March 14, 2023, Tools of the Trade Conference: The Way Forward

国立国会図書館
National Diet Library, Japan

New Features of Digital Services at the NDL

Onuma, Tahee

Assistant Director, Digital Information Planning Division

National Diet Library, Japan

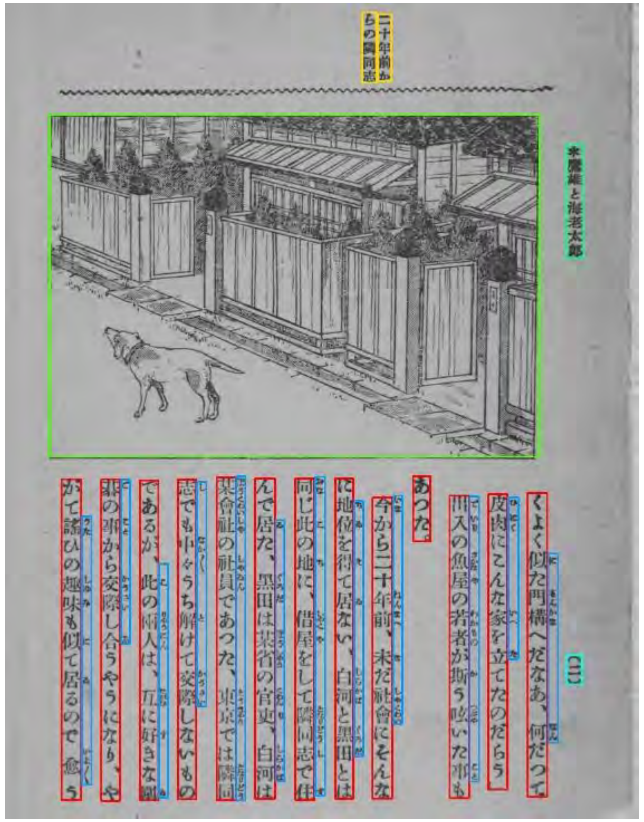
lab@ndl.go.jp

Impact of Machine Learning Technology

- The development of machine learning/AI technology has greatly expanded possibilities for information exploration.
  - Data mining
  - Natural language processing
  - Computer vision

  ...

- Significant recent advances at the NDL:
  - Services using **full-text data**
  - **Image search** functionality



二十年前からの隣同志

＊鷹雄と海老太郎

（三）

くよく似た門構へだなあ、何だつて皮肉にこんな家を立てたのだらう出入の魚屋の若者が斯う呟いた事もあつた。

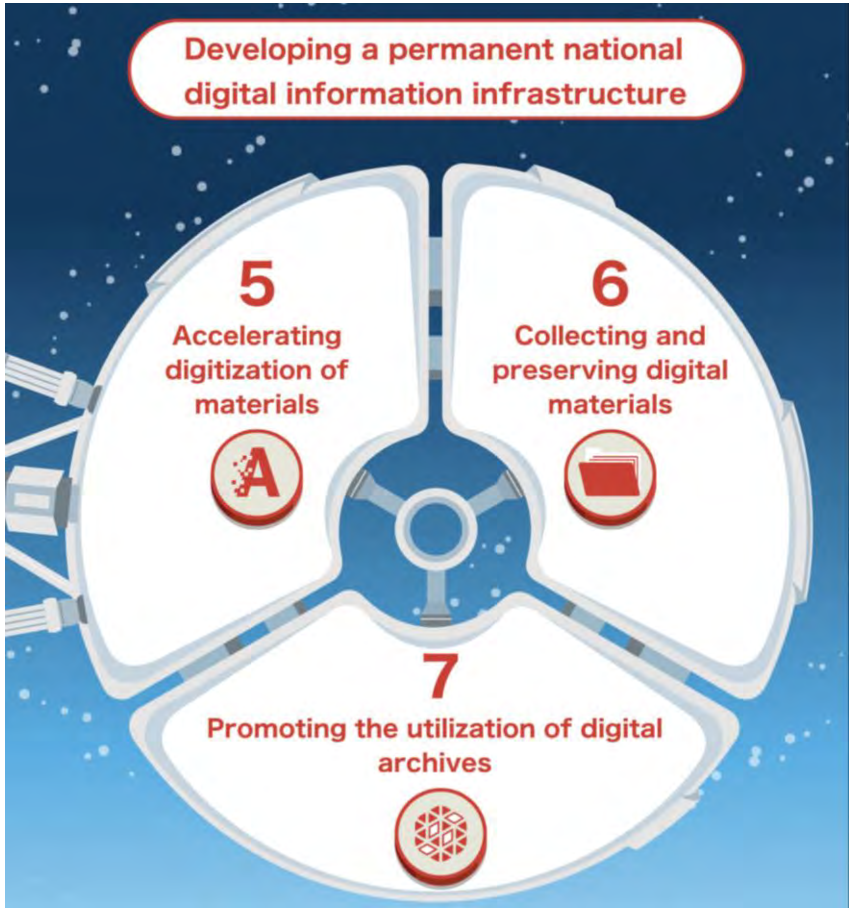
今から二十年前、未だ社會にそんなに地位を得て居ない、白河と黒田とは同じ此の地に、借屋をして隣同志で住んで居た。黒田は某省の官吏、白河は某會社の社員であつた、東京では隣同志でも中々うち解けて交際しないものであるが、此の兩人は、互に好きな園藝の事から交際し合うやうになり、やがて謠ひの趣味も似て居るので愈う

1. Use of Full-Text Data

- *Vision 2021-2025*

  Initiative 5: "**Accelerating digitization of materials**," which also includes full-text conversion

- Benefits of full-text data
  - For search purposes
  - Datasets for machine learning


Developing a permanent national digital information infrastructure
5
Accelerating digitization of materials
6
Collecting and preserving digital materials
7
Promoting the utilization of digital archives

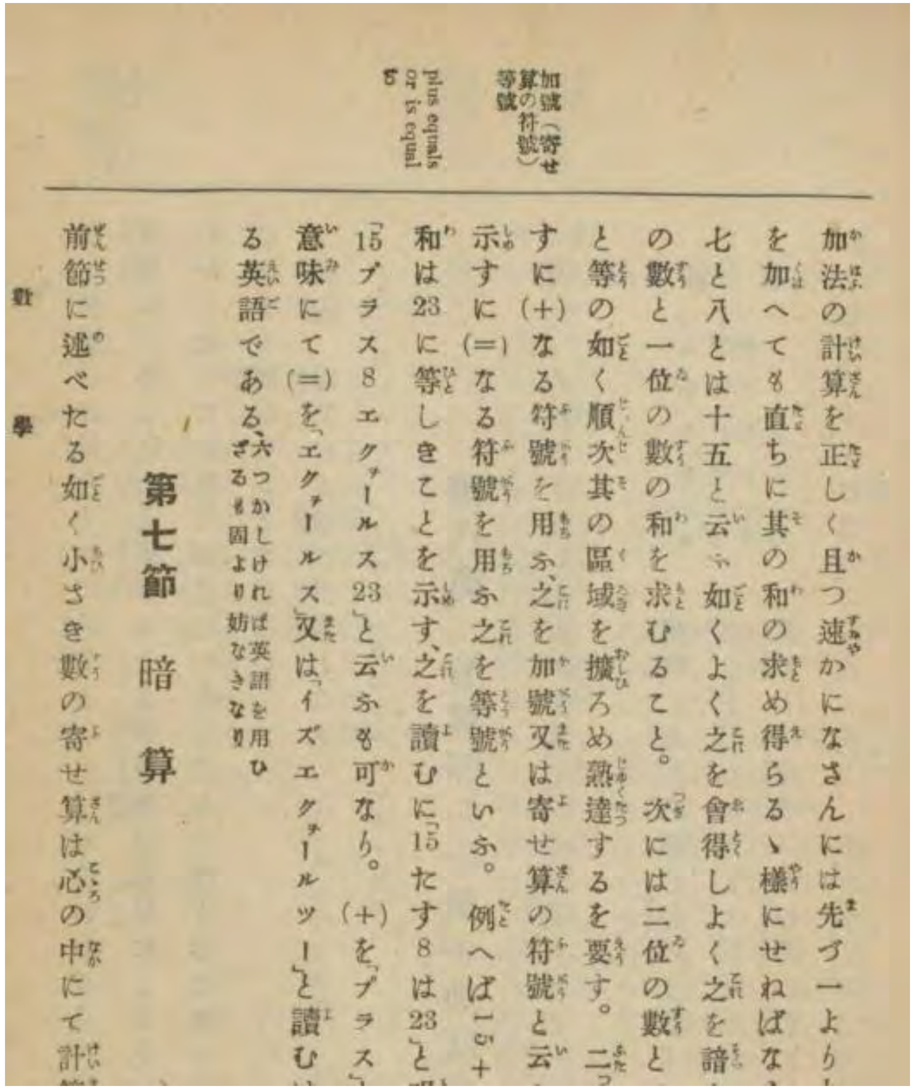
Challenges in OCR for Modern Japanese Materials

1. **Complexity of writing system**
   - **Kanji (漢字), hiragana (ひらがな), katakana (カタカナ)**
   - Latin alphabet and other signs (abc, 123, #$%&...)

2. **Mixture of reading direction**
   - Vertical or horizontal, sometimes mixed (!)

3. **Some typographic features**
   - Ruby characters (ルビ)
   - Headnotes (頭注), split annotations (割注)

…



加號（寄せ算の符號）等號

plus equals
or is equal
to

加法の計算を正しく且つ速かになさんには先づ一より
を加へても直ちに其の和の求め得らるゝ樣にせねばな
七と八とは十五と云ふ如くよく之を會得しよく之を諳
の數と一位の數の和を求むること。次には二位の數と
と等の如く順次其の區域を擴ろめ熟達するを要す。二つ
すに(+)なる符號を用ふ、之を加號又は寄せ算の符號と云
示すに(=)なる符號を用ふ之を等號といふ。例へば15+
和は23に等しきことを示す、之を讀むに「15たす8は23」と
「15プラス8エクヮールス23」と云ふも可なり。(+)を「プラス」
意味にて(=)を「エクヮールス」又は「イズエクヮールツー」と讀む
る英語である、六つかしければ英語を用ひざるも固より妨なきなり

第七節　暗算

前節に述べたる如く小さき數の寄せ算は心の中にて計

數學

https://doi.org/10.11501/1182117

OCR-related Projects in FY2021: Overview

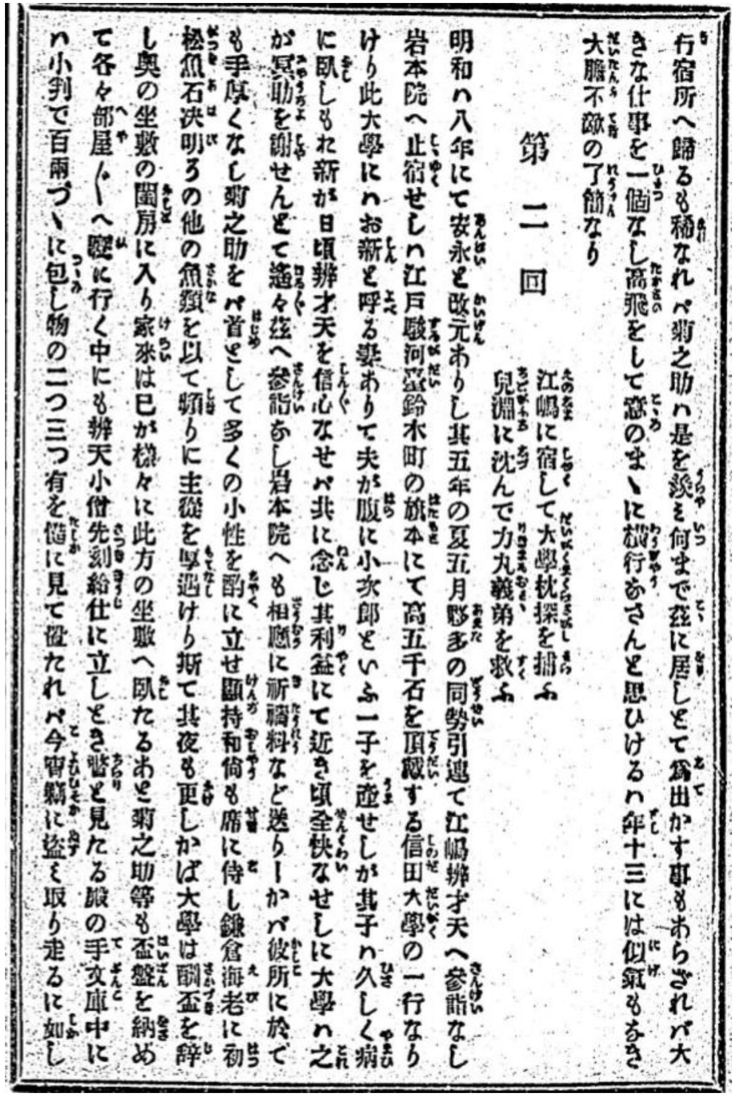
- **Project 1: Mass conversion from digitized images into text data**
  - Target: **2.47 million items** (223 million images)

    = almost **all** documents that had been digitized as of 2020

- **Project 2: Development of AI-OCR software for Japanese materials**
  - To be used for the text conversion of materials digitized after 2021
  - To be made freely available as open-source software (**OSS**)

OCR-related Projects in FY2021: Challenges

- How to generate full-text data for massive amounts of digitized images (***cost***, ***time***)
- How to deal with characteristics of and variety in pre-WW II (1868-1945) Japanese-language materials, such as outdated character forms or fonts (***quality***)



行宿所へ歸るも稀なれバ菊之助ハ是を羨ミ何まで茲に居しとて露出かす事もあらざれバ大きな仕事を一個なし高飛をして恣のまゝに横行あさんと思ひけるハ年十三には似氣もなき大膽不敵の了簡なり

第二回

江嶋に宿して大學枕探を捕ふ
兒淵に沈んで力丸義弟を救ふ

明和ハ八年にて安永と改元ありし其五年の夏五月夥多の同勢引連て江嶋辨才天へ參詣なし岩本院へ止宿せしハ江戸駿河臺鈴木町の旗本にて高五千石を頂戴する信田大學の一行なりけり此大學にハお新と呼る妻ありて夫が腹に小次郎といふ一子を産せしが其子ハ久しく病に臥しもれ新が日頃辨才天を信心なせバ共に念じ其利益にて近き頃全快なせしに大學ハ之が冥助を謝せんとて遙々茲へ參詣あし岩本院へも相應に祈禱料など送りしかバ彼所に於ても手厚くなし菊之助を首として多くの小性を酌に立せ顯持和尙も席に侍し鎌倉海老に初松魚石決明ろの他の魚類を以て頻りに主從を學遁けり斯て其夜も更しかば大學は酬盃を辭し奥の坐敷の閨房に入り家來は巳が樣々に此方の坐敷へ臥たるあと菊之助等も盃盤を納めて各々部屋〳〵へ寢に行く中にも辨天小僧先刻給仕に立しとき瞥と見たる殿の手文庫中にハ小判で百兩づゝに包し物の二つ三つ有を慥に見て從たれバ今宵竊に盜ミ取り走るに如し

https://doi.org/10.11501/881298

- Development of an AI-OCR model optimized for digitized materials of the NDL

OCR-related Projects in FY2021: Output

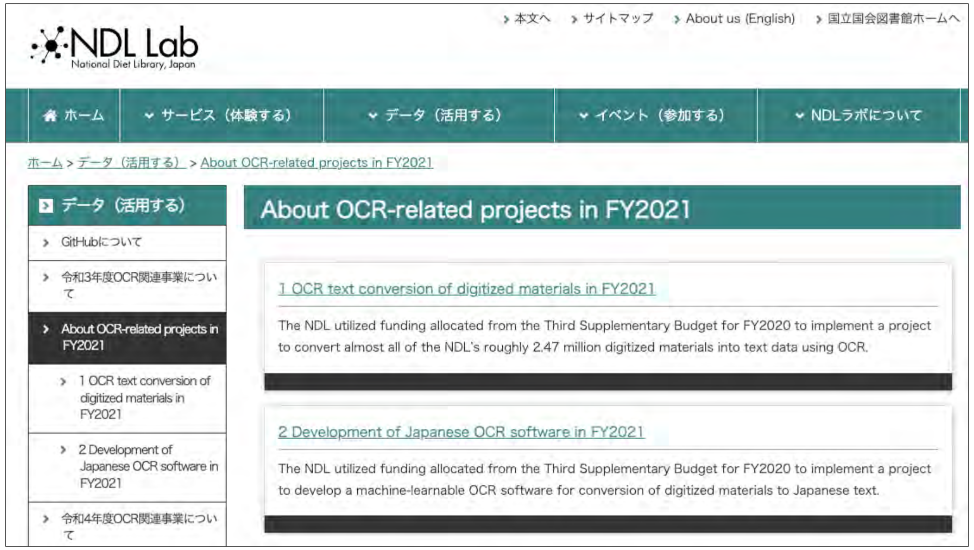
1. A huge amount of full-text data

- Books, periodicals and other digitized collections (uncorrected)
- Integrated into **NDL Digital Collections** for full-text search

2. AI-OCR software NDLOCR

- Customizable and trainable OSS
- Freely available on GitHub (CC-BY)

For more details of these OCR projects, please see:
https://lab.ndl.go.jp/data_set/ocr_en/


本文へ
サイトマップ
About us (English)
国立国会図書館ホームへ
NDL Lab
National Diet Library, Japan
ホーム
サービス（体験する）
データ（活用する）
イベント（参加する）
NDLラボについて
ホーム > データ（活用する） > About OCR-related projects in FY2021
データ（活用する）
GitHubについて
令和3年度OCR関連事業について
About OCR-related projects in FY2021
1 OCR text conversion of digitized materials in FY2021
2 Development of Japanese OCR software in FY2021
令和4年度OCR関連事業について

About OCR-related projects in FY2021

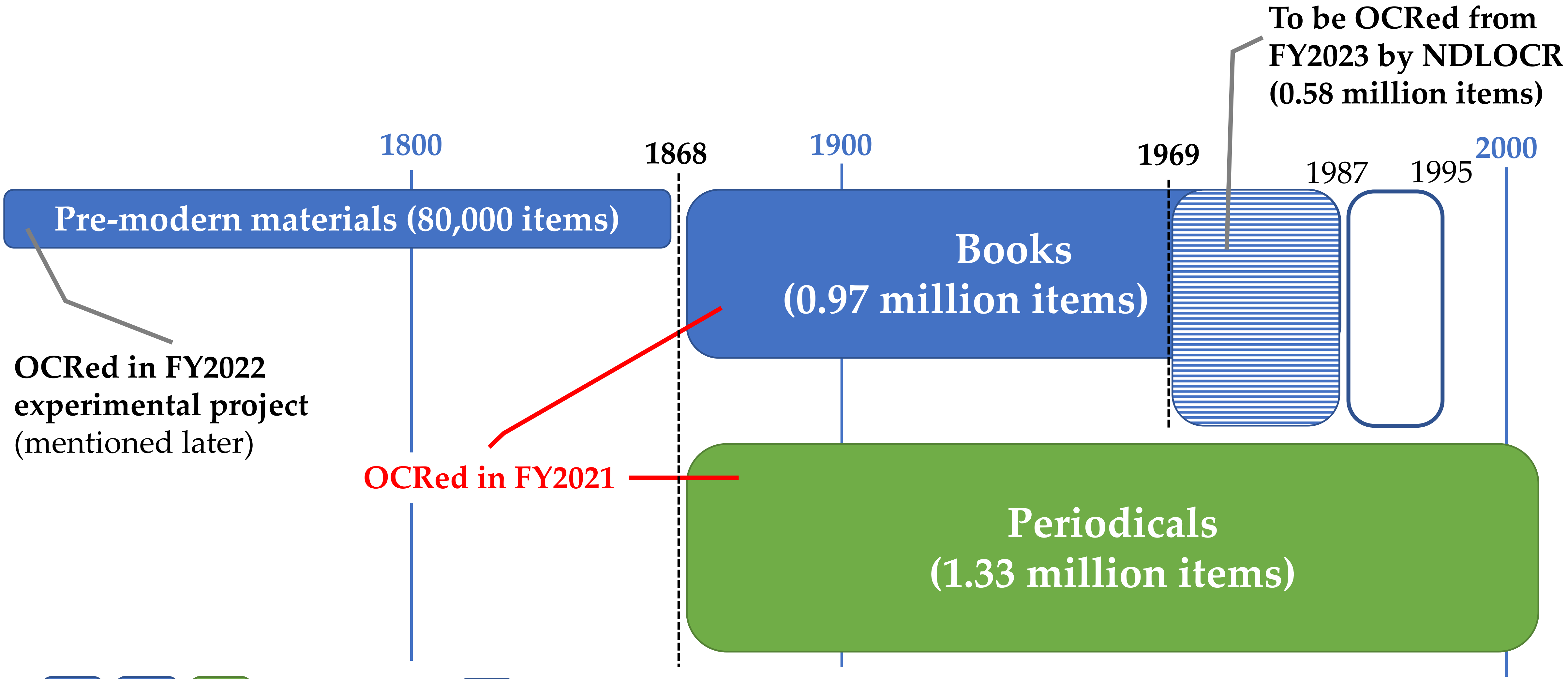
1 OCR text conversion of digitized materials in FY2021

The NDL utilized funding allocated from the Third Supplementary Budget for FY2020 to implement a project to convert almost all of the NDL's roughly 2.47 million digitized materials into text data using OCR.

2 Development of Japanese OCR software in FY2021

The NDL utilized funding allocated from the Third Supplementary Budget for FY2020 to implement a project to develop a machine-learnable OCR software for conversion of digitized materials to Japanese text.

Summary: Digitized and OCRed Materials at the NDL



Pre-modern materials (80,000 items)

Books
(0.97 million items)

Periodicals
(1.33 million items)

To be OCRed from
FY2023 by NDLOCR
(0.58 million items)

OCRed in FY2022
experimental project
(mentioned later)

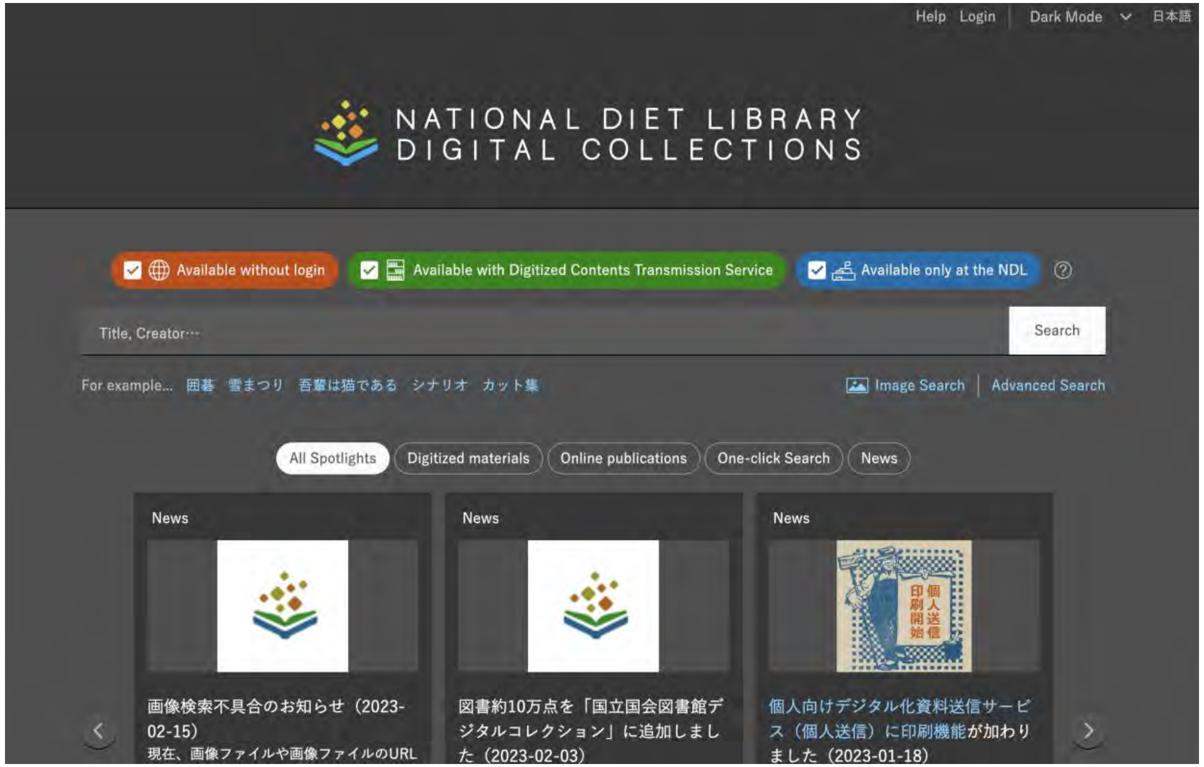
OCRed in FY2021

▇ ▤ 🟩 : Digitized □ : Prepared for digitization from FY2023

NDL Digital Collections: Overview

- Platform for digital materials collected by the NDL
  - Digitized documents (3.33 million items)
  - Online publications (1.46 million items)
  - Doctoral dissertations in digital format (0.1 million items)

- 3 levels of access:
  - Available online (mostly PD)
  - Available through the Digitized Contents Transmission Service (see next slide)
  - On-premises only



画像検索不具合のお知らせ（2023-02-15）
現在、画像ファイルや画像ファイルのURL

図書約10万点を「国立国会図書館デジタルコレクション」に追加しました（2023-02-03）

個人向けデジタル化資料送信サービス（個人送信）に印刷機能が加わりました（2023-01-18）

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

Digitized Contents Transmission Service

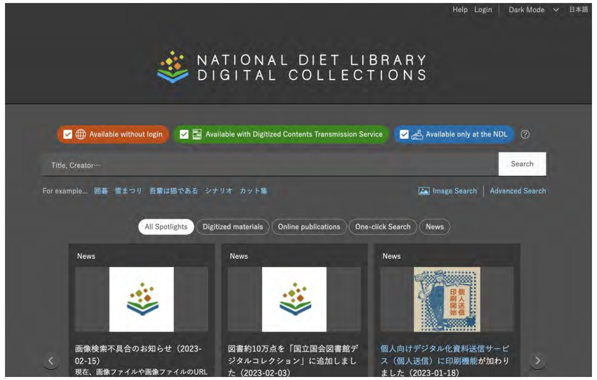
- Provides online access to digitized out-of-print or otherwise difficult-to-obtain materials (including those under copyright protection)

i. For libraries (January 2014-)

- In partner libraries, patrons can access materials at terminals on premises

ii. For individuals (May 2022-)

- Official registered users (Japanese residents only) can access them via internet



Help Login | Dark Mode 日本語

NATIONAL DIET LIBRARY DIGITAL COLLECTIONS

☑ Available without login ☑ Available with Digitized Contents Transmission Service ☑ Available only at the NDL

Title, Creator… Search

For example... 囲碁 雪まつり 吾輩は猫である シナリオ カット集 Image Search | Advanced Search

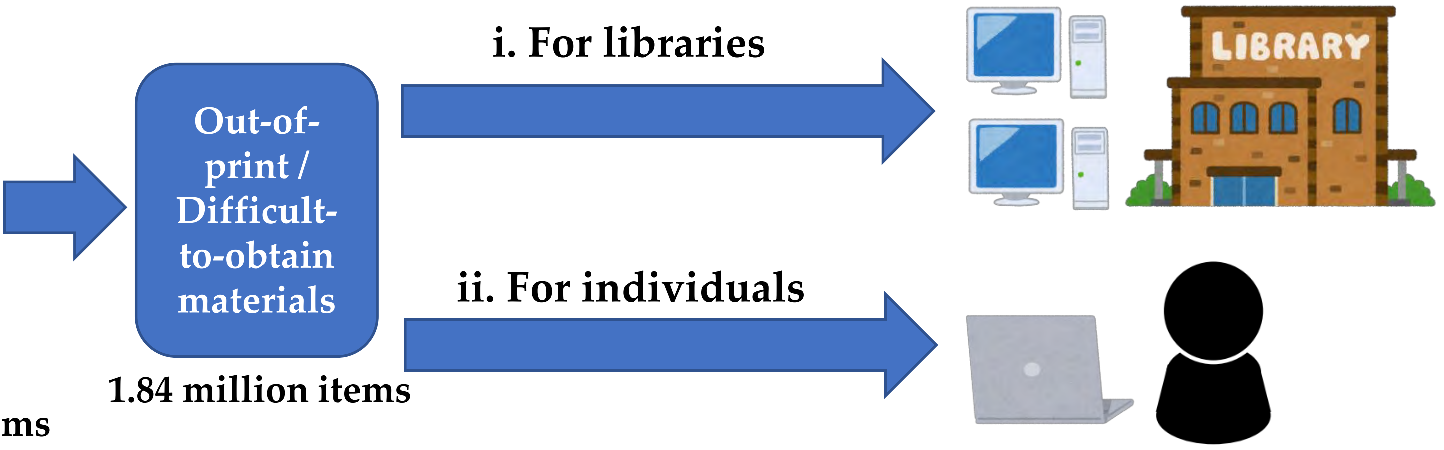
All Spotlights Digitized materials Online publications One-click Search News

News
画像検索不具合のお知らせ（2023-02-15）
現在、画像ファイルや画像ファイルのURL

News
図書約10万点を「国立国会図書館デジタルコレクション」に追加しました（2023-02-03）

News
個人向けデジタル化資料送信サービス（個人送信）に印刷機能が加わりました（2023-01-18）

Digitized materials: 3.33 million items



Out-of-print / Difficult-to-obtain materials

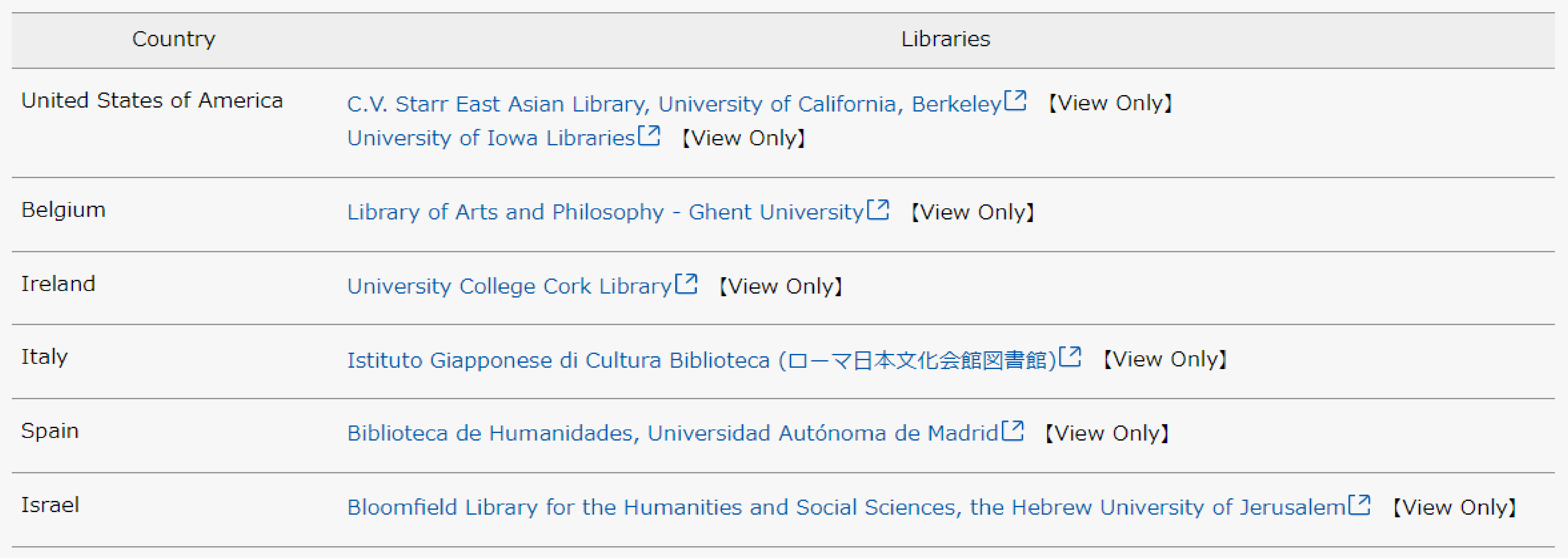
1.84 million items

i. For libraries

ii. For individuals

Digitized Contents Transmission Service (for Libraries)

- **List of overseas partner libraries** (as of March 1, 2023)

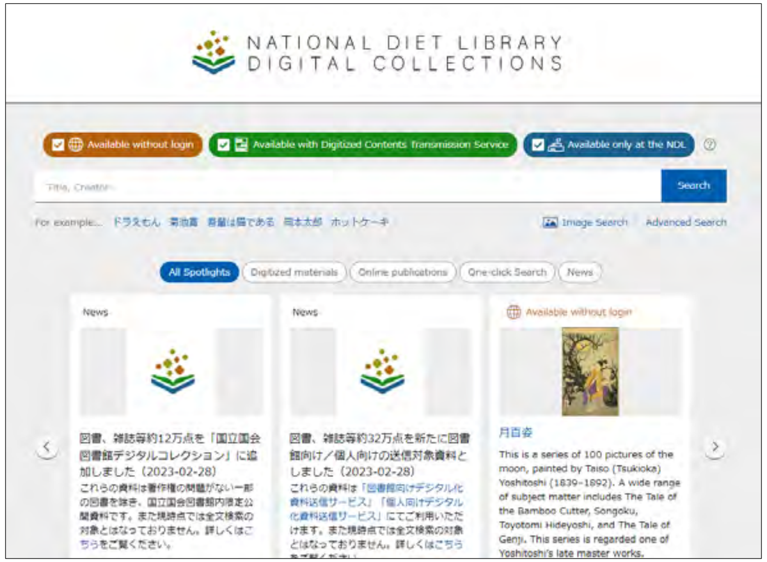


Country	Libraries
United States of America	C.V. Starr East Asian Library, University of California, Berkeley 【View Only】 University of Iowa Libraries 【View Only】
Belgium	Library of Arts and Philosophy - Ghent University 【View Only】
Ireland	University College Cork Library 【View Only】
Italy	Istituto Giapponese di Cultura Biblioteca (ローマ日本文化会館図書館) 【View Only】
Spain	Biblioteca de Humanidades, Universidad Autónoma de Madrid 【View Only】
Israel	Bloomfield Library for the Humanities and Social Sciences, the Hebrew University of Jerusalem 【View Only】

https://dl.ndl.go.jp/en/soshin_librarylist

NDL Digital Collections: Recent Progress

- Acceleration of digitization (as of February 2023)
  - Available online: 0.58 million items
  - Available through transmission service: 1.84 million items
  - On-premises: 0.91 million items
- Totally renewed in December 2022 with new functionalities:
  - **Full-text search**
  - **Image search**



NATIONAL DIET LIBRARY DIGITAL COLLECTIONS

☑ Available without login ☑ Available with Digitized Contents Transmission Service ☑ Available only at the NDL

News

図書、雑誌等約12万点を「国立国会図書館デジタルコレクション」に追加しました（2023-02-28）

これらの資料は著作権の問題がない一部の図書を除き、国立国会図書館内限定公開資料です。また現時点では全文検索の対象とはなっておりません。詳しくはこちらをご覧ください。

News

図書、雑誌等約32万点を新たに図書館向け／個人向けの送信対象資料としました（2023-02-28）

これらの資料は「図書館向けデジタル化資料送信サービス」「個人向けデジタル化資料送信サービス」にてご利用いただけます。また現時点では全文検索の対象とはなっておりません。詳しくはこちら

Available without login

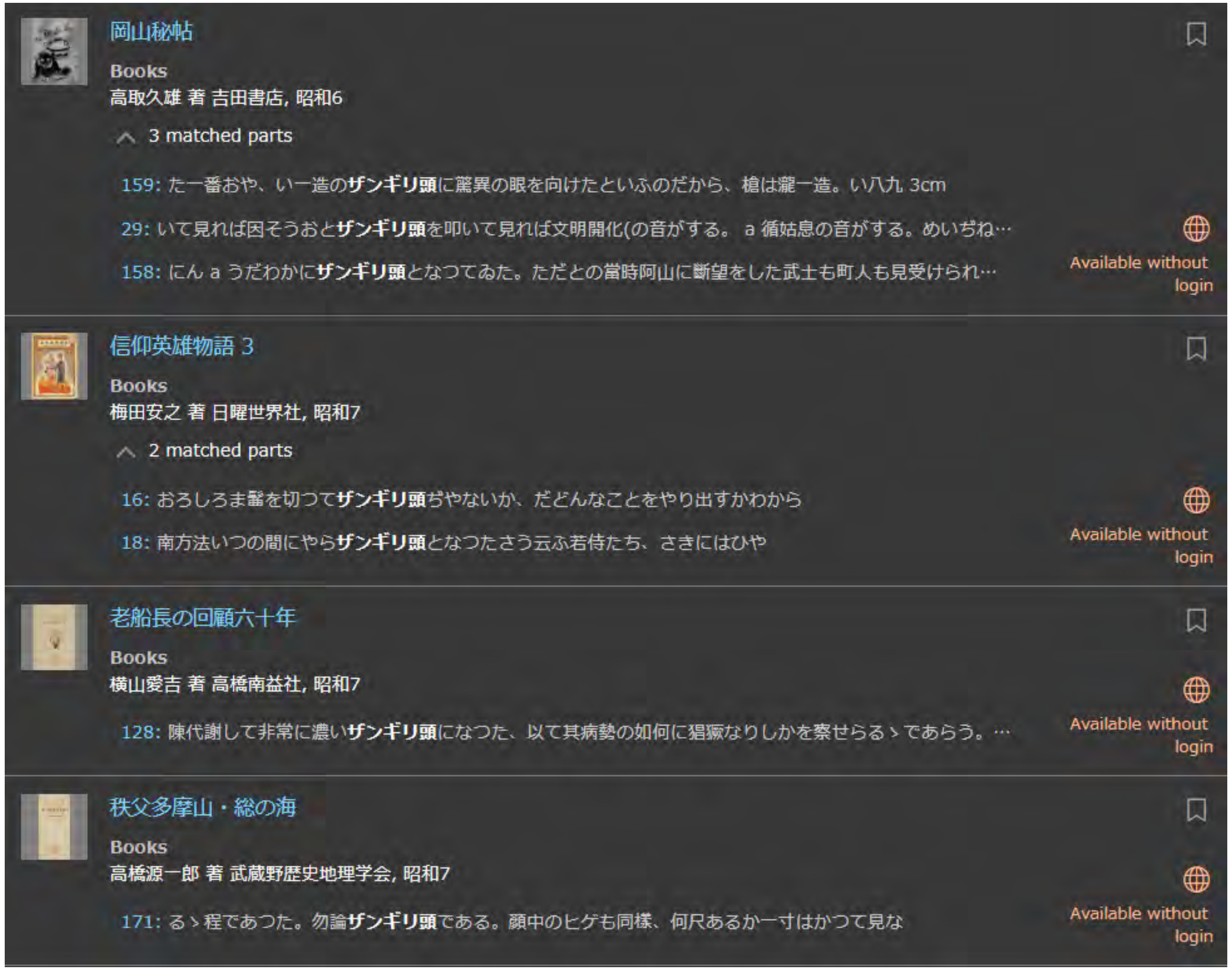
月百姿

This is a series of 100 pictures of the moon, painted by Taiso (Tsukioka) Yoshitoshi (1839-1892). A wide range of subject matter includes The Tale of the Bamboo Cutter, Songoku, Toyotomi Hideyoshi, and The Tale of Genji. This series is regarded one of Yoshitoshi's late master works.

cf.
"Digitizing Library Materials at the NDL (part 1)", *National Diet Library Newsletter*, (249).
"Renewal of the National Diet Library Digital Collections", ibid.

NDL Digital Collections: Full-Text Search

- Text data produced in FY 2021 has been integrated. (2.47 million items)
- In the search results page, query keywords and their surrounding text are displayed as snippets.
- This functionality has led to a significant improvement in possibilities for information seeking. (e.g. first appearance of a specific term in the corpus)



岡山秘帖

Books

高取久雄 著 吉田書店, 昭和6

3 matched parts

159: た一番おや、い一造の**ザンギリ頭**に驚異の眼を向けたといふのだから、槍は瀧一造。い八九 3cm

29: いて見れば因そうおと**ザンギリ頭**を叩いて見れば文明開化(の音がする。 a 循姑息の音がする。めいぢね…

158: にん a うだわかに**ザンギリ頭**となつてゐた。ただとの當時阿山に斷望をした武士も町人も見受けられ…

Available without login

信仰英雄物語 3

Books

梅田安之 著 日曜世界社, 昭和7

2 matched parts

16: おろしろま髷を切つて**ザンギリ頭**ぢやないか、だどんなことをやり出すかわから

18: 南方法いつの間にやら**ザンギリ頭**となつたさう云ふ若侍たち、さきにはひや

Available without login

老船長の回顧六十年

Books

横山愛吉 著 高橋南益社, 昭和7

128: 陳代謝して非常に濃い**ザンギリ頭**になつた、以て其病勢の如何に猖獗なりしかを察せらる〻であらう。…

Available without login

秩父多摩山・総の海

Books

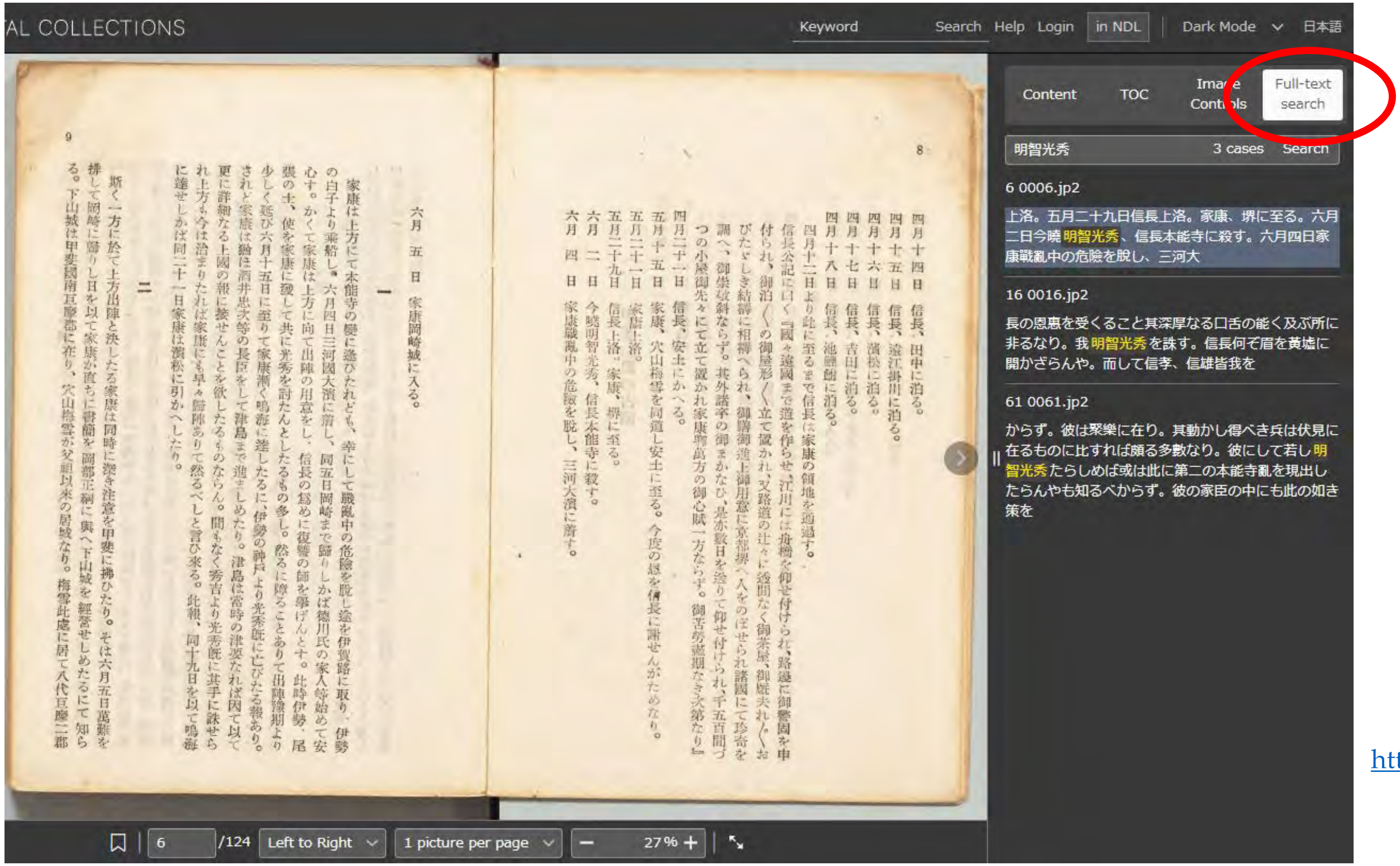
高橋源一郎 著 武蔵野歴史地理学会, 昭和7

171: る〻程であつた。勿論**ザンギリ頭**である。顔中のヒゲも同様、何尺あるか一寸はかつて見な

Available without login

NDL Digital Collections: Full-Text Search

- Full-text search within a document is also available.
- Snippets are displayed for PD documents (as shown below).



8

四月十四日　信長、田中に泊る。
四月十五日　信長、遠江掛川に泊る。
四月十六日　信長、濱松に泊る。
四月十七日　信長、吉田に泊る。
四月十八日　信長、池鯉鮒に泊る。

四月十二日より此に至るまで信長は家康の領地を通過す。信長公記に曰く『國々遠國まで道を作らせ、江川には舟橋を仰せ付けられ、路邊に御警固を申付られ、御泊〱の御屋形〱立て置かれ、又路道の辻々に透間なく御茶屋、御厩夫れ〱おびたゞしき結構に相構へられ、御賄御進上御用意に京都堺へ人をのぼせられ諸國にて珍奇を調へ、御崇敬斜ならず。其外諸卒の御まかなひ、是亦數日を逐りて仰せ付けられ、千五百間づつの小屋御先々にて立て置かれ家康卿萬方の御心賦一方ならず。御苦勞盡期なき次第なり』

四月二十一日　信長、安土にかへる。
五月十五日　家康、穴山梅雪を同道し安土に至る。今度の懇を信長に謝せんがためなり。
五月二十一日　家康上洛。
五月二十九日　信長上洛。家康、堺に至る。
六月　二　日　今曉明智光秀、信長本能寺に殺す。
六月　四　日　家康戰亂中の危險を脫し、三河大濱に着す。

9

六月　五　日　家康岡崎城に入る。

一

家康は上方にて本能寺の變に遭ひたれども、幸にして戰亂中の危險を脫し途を伊賀路に取り、伊勢の白子より乘船し、六月四日三河國大濱に着し、同五日岡崎まで歸りしかば德川氏の家人等始めて安心す。かくて家康は上方に向て出陣の用意をし、信長の爲めに復讐の師を擧げんとす。此時伊勢、尾張の士、使を家康に發して共に光秀を討たんとしたるもの多し。然るに障ることありて出陣猶豫期より少しく延び六月十五日に至りて家康漸く鳴海に達したるに、伊勢の神戸より光秀既に亡びたる報あり。されど家康は猶は酒井忠次等の長臣をして津島まで進ましめたり。津島は當時の津要たれば因て以て更に詳細なる上國の報に接せんことを欲したるものならん。間もなく秀吉より光秀既に其手に誅せられ上方も今は治まりたれば家康にも早々歸陣ありて然るべしと言ひ來る。此報、同十九日を以て鳴海に達せしかば同二十一日家康は濱松に引かへしたり。

二

斯く一方に於て上方出陣と決したる家康は同時に深き注意を甲斐に拂ひたり。それは六月五日萬難を排して岡崎に歸りし日を以て家康が直ちに書簡を岡部正綱に與へ下山城を經營せしめたるにて知らる。下山城は甲斐國南巨摩郡に在り、穴山梅雪が父祖以來の居城なり。梅雪此處に居て八代巨摩二郡

明智光秀　3 cases　Search

6 0006.jp2

上洛。五月二十九日信長上洛。家康、堺に至る。六月二日今曉明智光秀、信長本能寺に殺す。六月四日家康戰亂中の危險を脫し、三河大

16 0016.jp2

長の恩惠を受くること其深厚なる口舌の能く及ぶ所に非るなり。我明智光秀を誅す。信長何ぞ眉を黄壚に開かざらんや。而して信孝、信雄皆我を

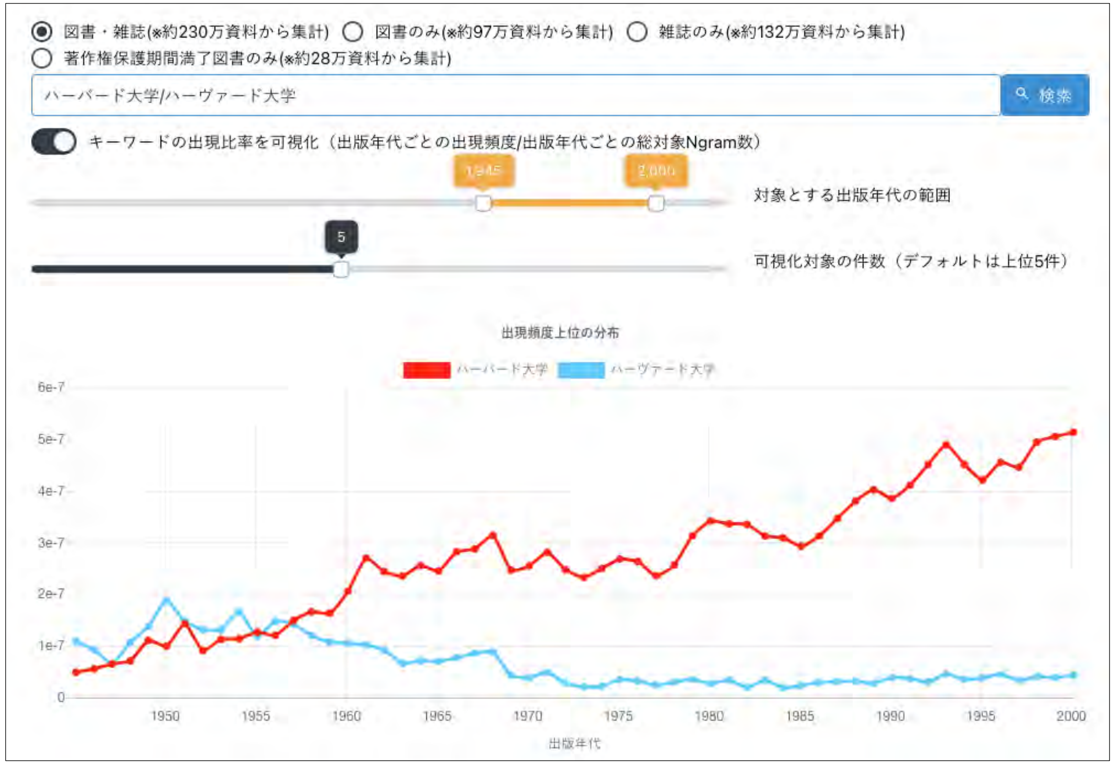
61 0061.jp2

からず。彼は聚樂に在り。其動かし得べき兵は伏見に在るものに比すれば頗る多數なり。彼にして若し明智光秀たらしめば或は此に第二の本能寺亂を現出したらんやも知るべからず。彼の家臣の中にも此の如き策を

https://doi.org/10.11501/1907107

NDL Ngram Viewer

- Experimental analytic tool for the corpus of digitized **books** and **periodicals** (2.3 million items)
- Visualizes frequency of query keywords by publication year
- Useful features:
  - Supports **regular expression** search
  - Target corpus can be selected (only books, periodicals, or PD books…)



● 図書・雑誌(※約230万資料から集計) ○ 図書のみ(※約97万資料から集計) ○ 雑誌のみ(※約132万資料から集計)
○ 著作権保護期間満了図書のみ(※約28万資料から集計)

ハーバード大学/ハーヴァード大学

キーワードの出現比率を可視化（出版年代ごとの出現頻度/出版年代ごとの総対象Ngram数）

対象とする出版年代の範囲: 1,945 – 2,000

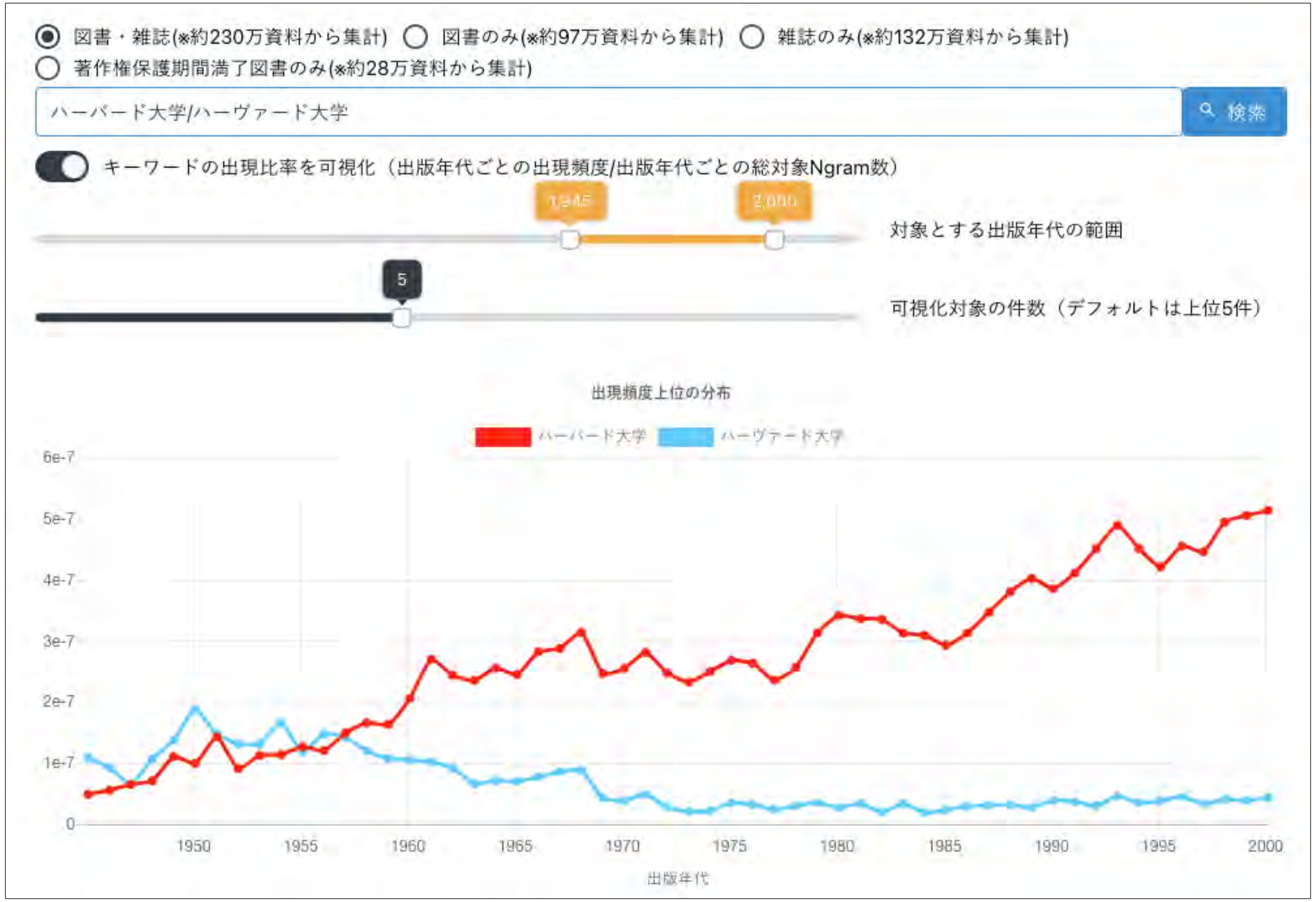
可視化対象の件数（デフォルトは上位5件）: 5

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

For more information, see: https://lab.ndl.go.jp/service/ngramviewer/en/

NDL Ngram Viewer

- Example: Comparison of two variants for “Harvard University”
  in Japanese writing (“ハーバード大学” or “ハーヴァード大学”)



● 図書・雑誌(※約230万資料から集計) ○ 図書のみ(※約97万資料から集計) ○ 雑誌のみ(※約132万資料から集計)
○ 著作権保護期間満了図書のみ(※約28万資料から集計)

ハーバード大学/ハーヴァード大学

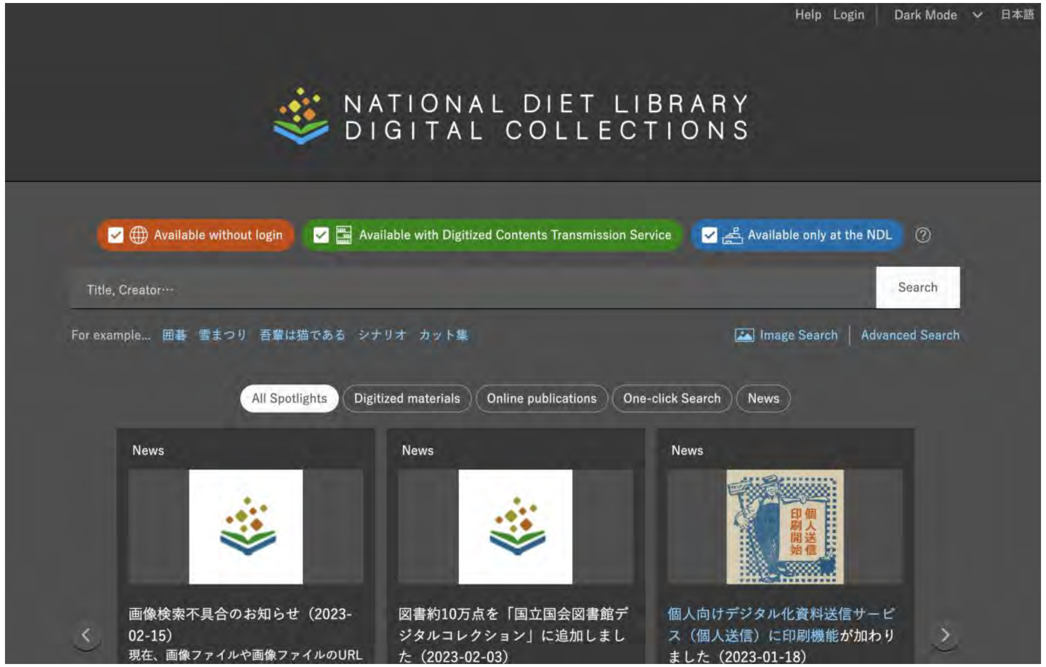
キーワードの出現比率を可視化（出版年代ごとの出現頻度/出版年代ごとの総対象Ngram数）

対象とする出版年代の範囲: 1,945 – 2,000

可視化対象の件数（デフォルトは上位5件）: 5

2. Use of Image Data

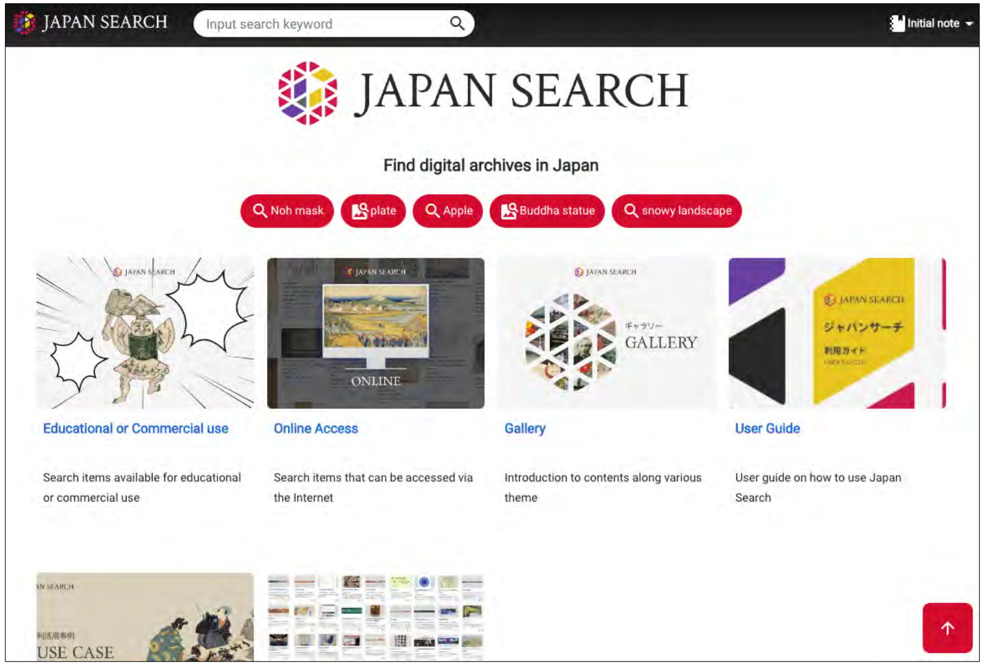
- Image-based search services offered by the NDL:
  - **NDL Digital Collections**
  - **Japan Search**


Help Login
Dark Mode
日本語
NATIONAL DIET LIBRARY DIGITAL COLLECTIONS
☑ Available without login
☑ Available with Digitized Contents Transmission Service
☑ Available only at the NDL
Title, Creator…
Search
For example... 囲碁 雪まつり 吾輩は猫である シナリオ カット集
Image Search
Advanced Search
All Spotlights
Digitized materials
Online publications
One-click Search
News
News
News
News
印刷開始
個人送信

画像検索不具合のお知らせ（2023-02-15）
現在、画像ファイルや画像ファイルのURL

図書約10万点を「国立国会図書館デジタルコレクション」に追加しました（2023-02-03）

個人向けデジタル化資料送信サービス（個人送信）に印刷機能が加わりました（2023-01-18）



Find digital archives in Japan

Educational or Commercial use
Search items available for educational or commercial use

Online Access
Search items that can be accessed via the Internet

Gallery
Introduction to contents along various theme

User Guide
User guide on how to use Japan Search

NDL Digital Collections: Image Search

- Image search functionality: released in December 2022
- Queries can be launched by:
  - Uploading an image file
  - Referring an image by URI

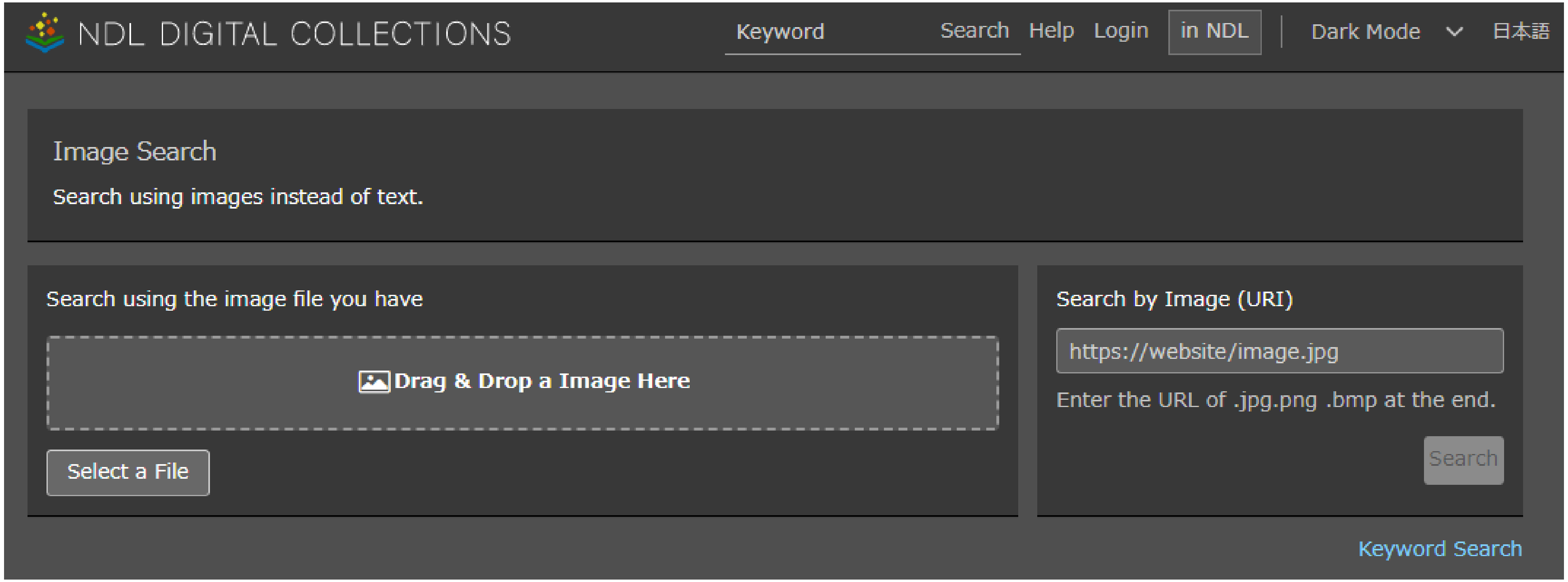


Image Search

Search using images instead of text.

Search using the image file you have

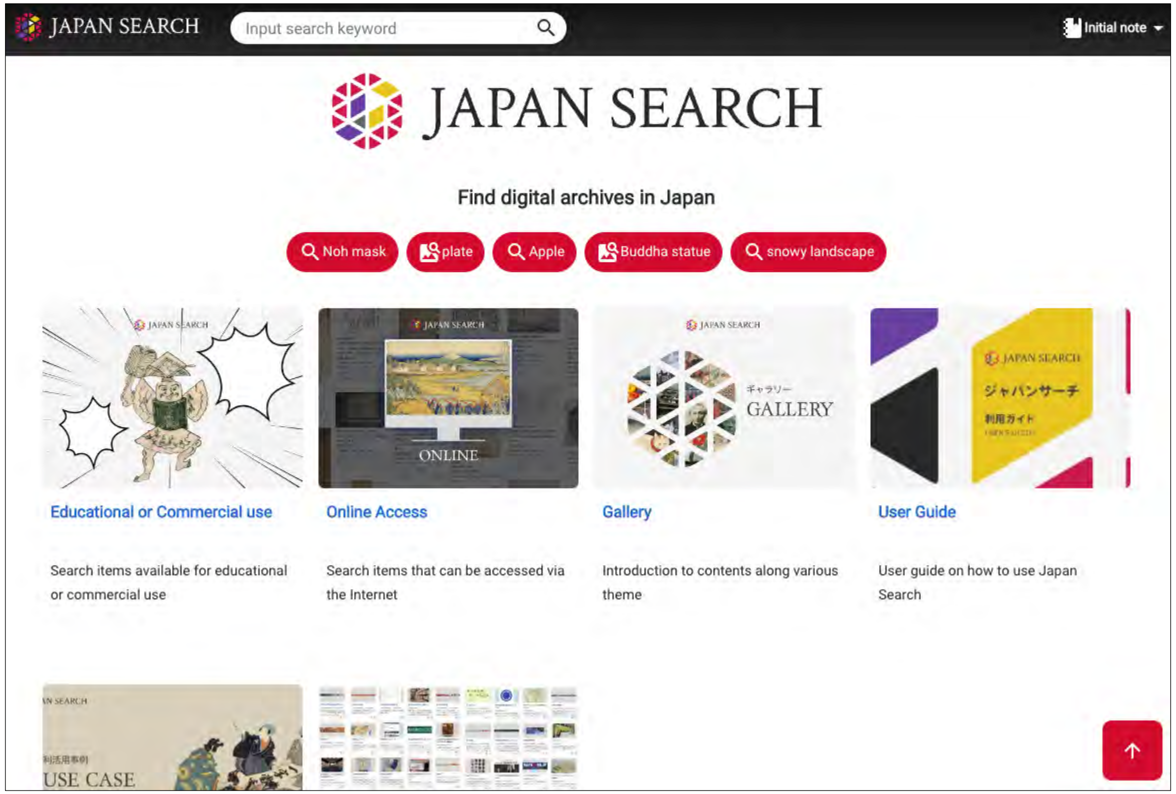
Search by Image (URI)

https://website/image.jpg

Enter the URL of .jpg.png .bmp at the end.

Japan Search: Overview

- Official release in August 2020
- National platform for promoting the use of content from Japan’s digital archives (such as GLAMs)
- Includes various types of digital materials (images, videos, sounds, 3D data…)
- 3 main functions: *Search*, *Galleries*, *API* (SPARQL)



Find digital archives in Japan

Educational or Commercial use

Search items available for educational or commercial use

Online Access

Search items that can be accessed via the Internet

Gallery

Introduction to contents along various theme

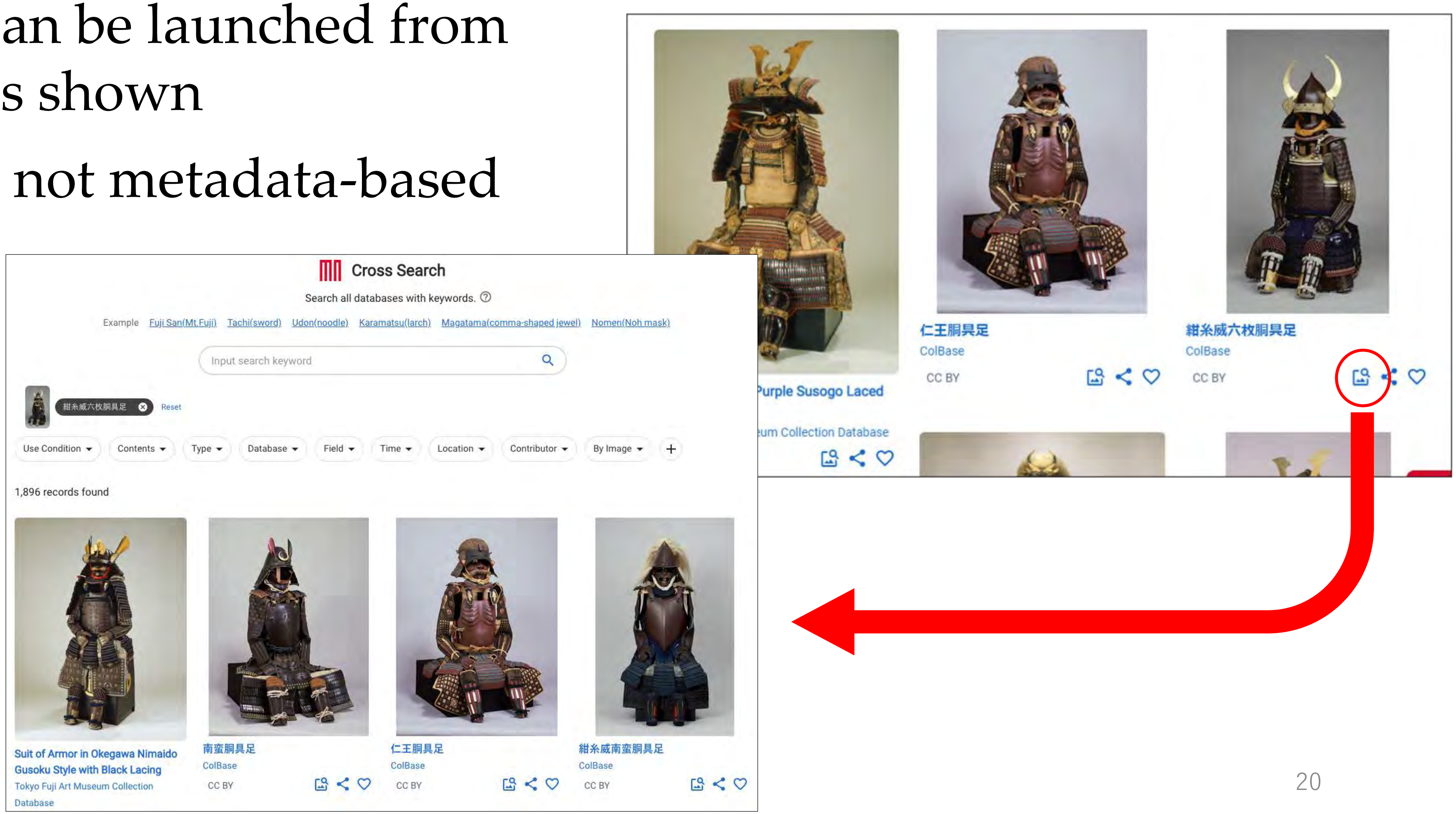
User Guide

User guide on how to use Japan Search

https://jpsearch.go.jp/

Japan Search: Image Search

- Image search can be launched from any thumbnails shown
- Feature-based, not metadata-based



Cross Search

Search all databases with keywords.

Example Fuji San(Mt.Fuji) Tachi(sword) Udon(noodle) Karamatsu(larch) Magatama(comma-shaped jewel) Nomen(Noh mask)

Input search keyword

紺糸威六枚胴具足 Reset

1,896 records found

Suit of Armor in Okegawa Nimaido Gusoku Style with Black Lacing
Tokyo Fuji Art Museum Collection Database

南蛮胴具足
ColBase
CC BY

仁王胴具足
ColBase
CC BY

紺糸威南蛮胴具足
ColBase
CC BY

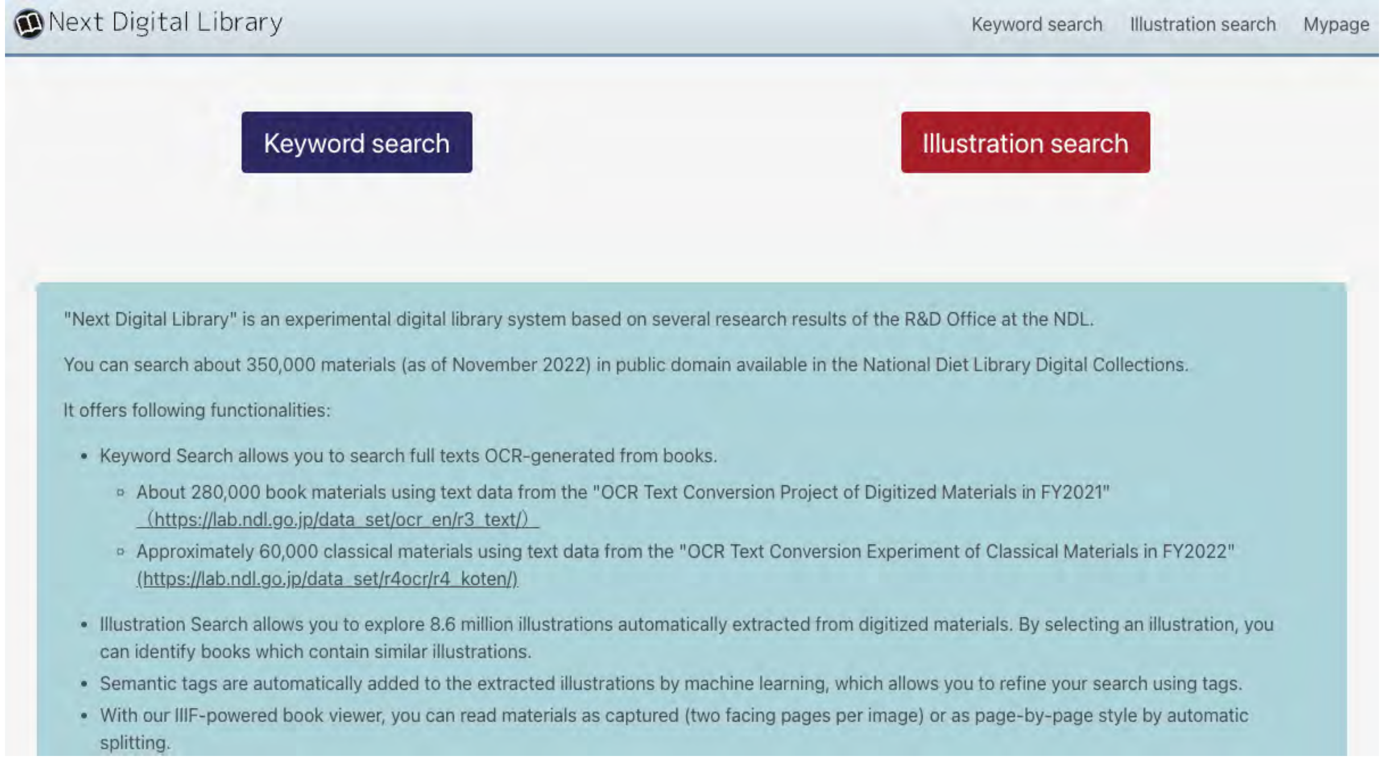
Purple Susogo Laced
um Collection Database

仁王胴具足
ColBase
CC BY

紺糸威六枚胴具足
ColBase
CC BY

Next Digital Library: Overview

- Experimental search/retrieval system for field trials of new functionalities developed by the R&D Office (March 2019-)
- Contains **books** and **pre-modern materials** in PD (350,000 in total)



"Next Digital Library" is an experimental digital library system based on several research results of the R&D Office at the NDL.

You can search about 350,000 materials (as of November 2022) in public domain available in the National Diet Library Digital Collections.

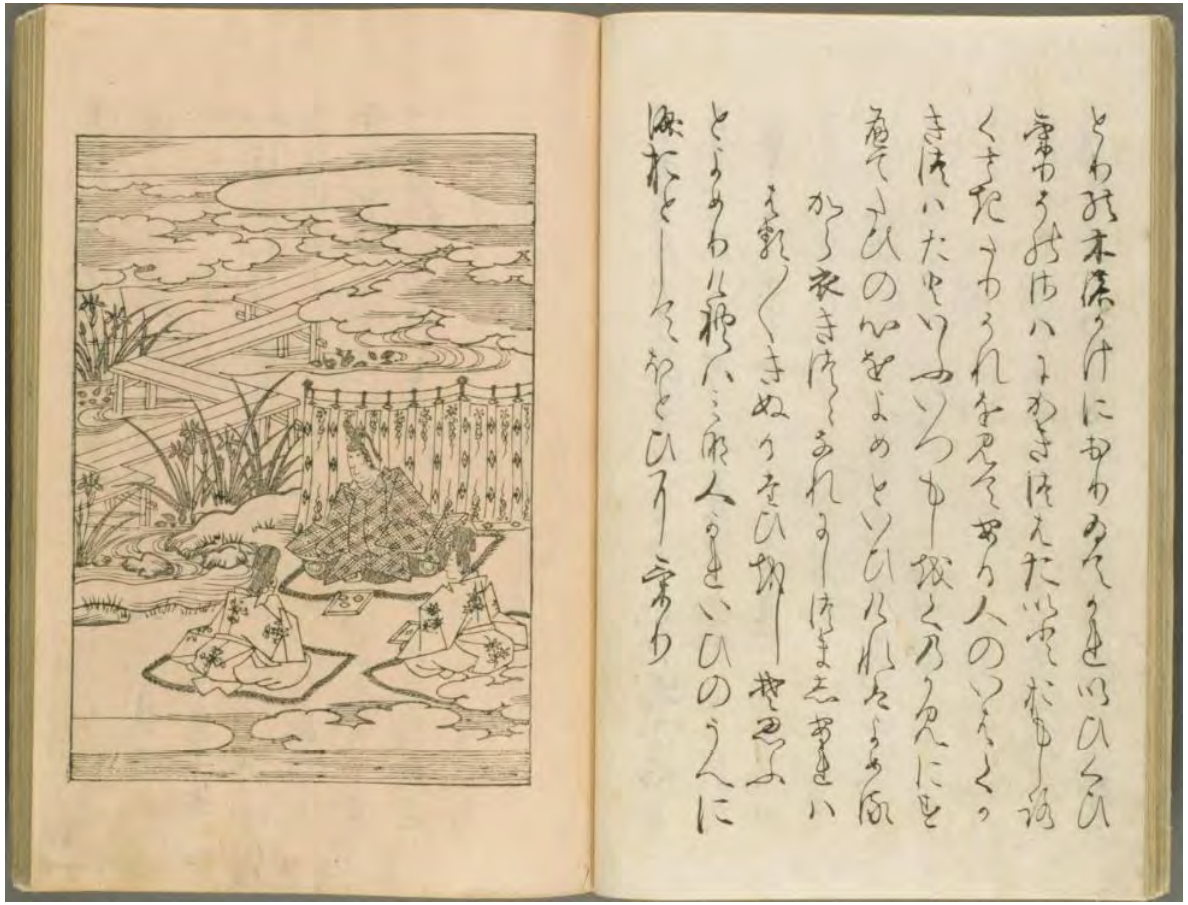
It offers following functionalities:

- Keyword Search allows you to search full texts OCR-generated from books.
  - About 280,000 book materials using text data from the "OCR Text Conversion Project of Digitized Materials in FY2021" （https://lab.ndl.go.jp/data_set/ocr_en/r3_text/）
  - Approximately 60,000 classical materials using text data from the "OCR Text Conversion Experiment of Classical Materials in FY2022" (https://lab.ndl.go.jp/data_set/r4ocr/r4_koten/)
- Illustration Search allows you to explore 8.6 million illustrations automatically extracted from digitized materials. By selecting an illustration, you can identify books which contain similar illustrations.
- Semantic tags are automatically added to the extracted illustrations by machine learning, which allows you to refine your search using tags.
- With our IIIF-powered book viewer, you can read materials as captured (two facing pages per image) or as page-by-page style by automatic splitting.

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

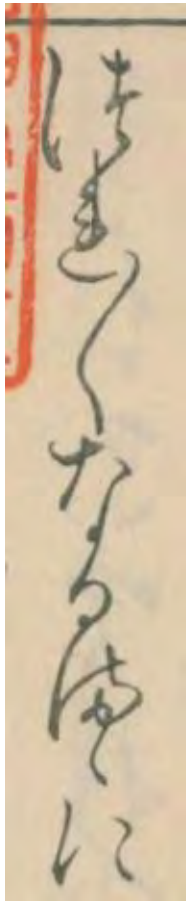
Challenges in OCR for Pre-modern Japanese Materials

Cursive characters (崩し字)



https://doi.org/10.11501/1287963

Variants of hiragana (変体仮名)

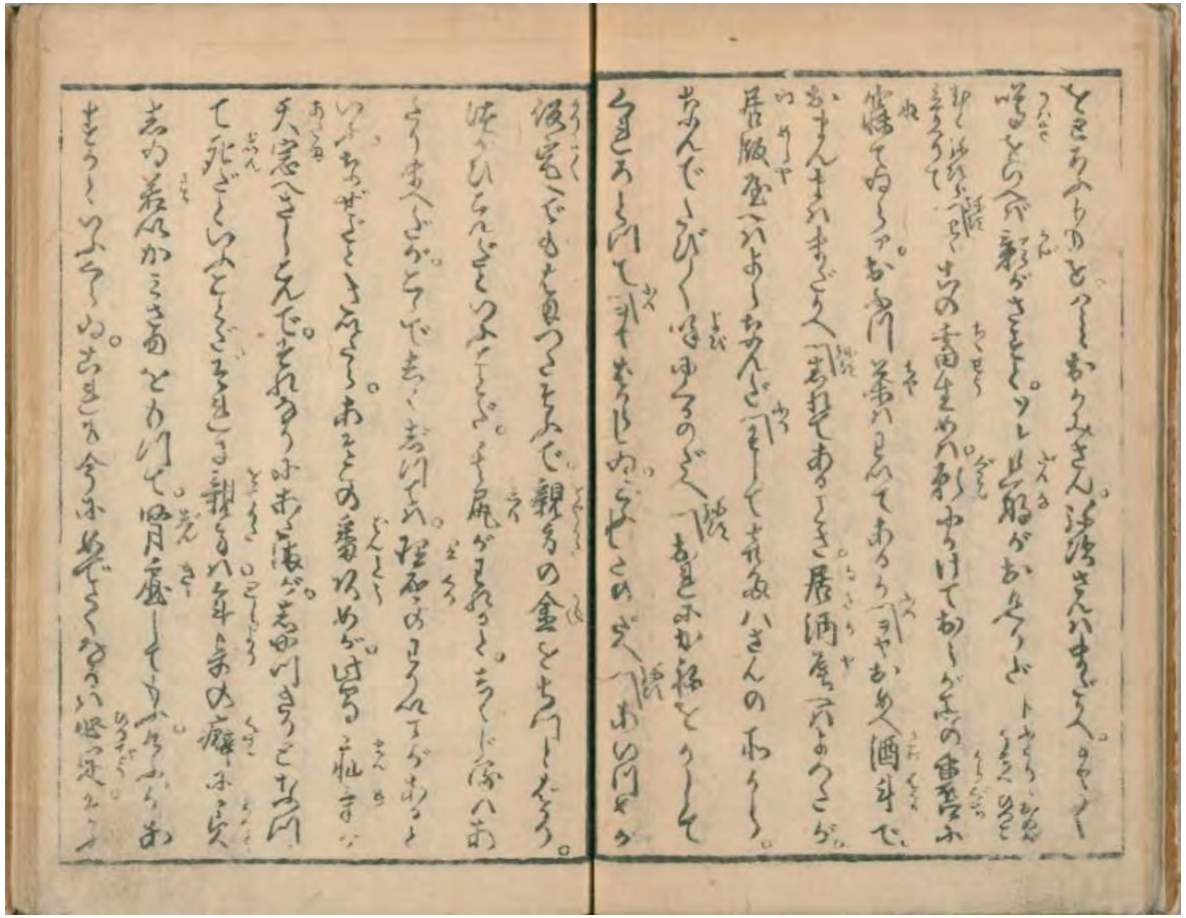


＝ つれづれなるままに

https://doi.org/10.11501/2534221

OCR Experiment for Pre-modern Materials

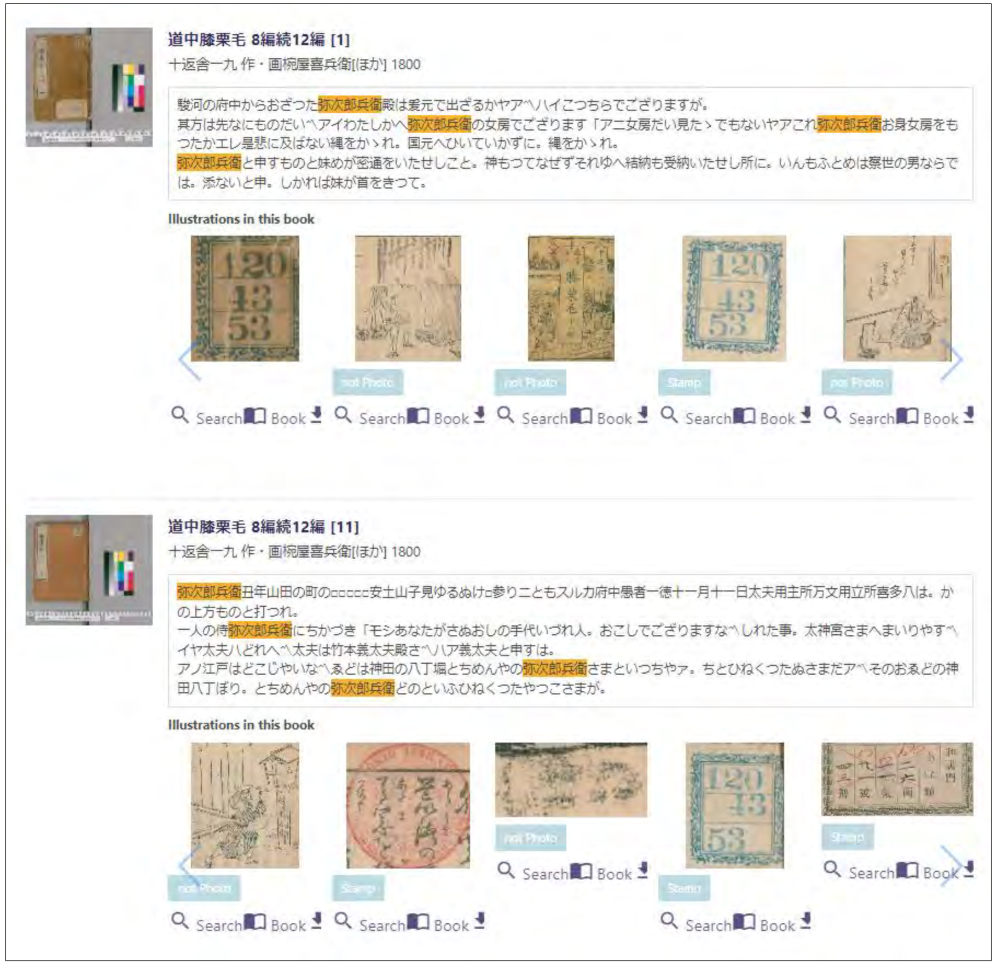
- In FY2022, the R&D Office developed in-house an AI-OCR software specialized in pre-modern materials (mostly before 1868) and generated full-text data.
- The full-text search for all digitized pre-modern materials (80,000 items) is available.



https://doi.org/10.11501/2558997

Next Digital Library: Full-Text Search

- In search results, snippets are displayed with query keywords highlighted.



道中膝栗毛 8編続12編 [1]

十返舎一九 作・画椀屋喜兵衛[ほか] 1800

駿河の府中からおざつた弥次郎兵衛殿は爰元で出ざるかヤア〽ハイこつちらでござりますが。
其方は先なにものだい〽アイわたしかへ弥次郎兵衛の女房でござります「アニ女房だい見たゝでもないヤアこれ弥次郎兵衛お身女房をもつたかエレ是悲に及ばない縄をかゝれ。国元へひいていかずに。縄をかゝれ。
弥次郎兵衛と申すものと妹めが密通をいたせしこと。神もつてなぜずそれゆへ結納も受納いたせし所に。いんもふとめは察世の男ならでは。添ないと申。しかれば妹が首をきつて。

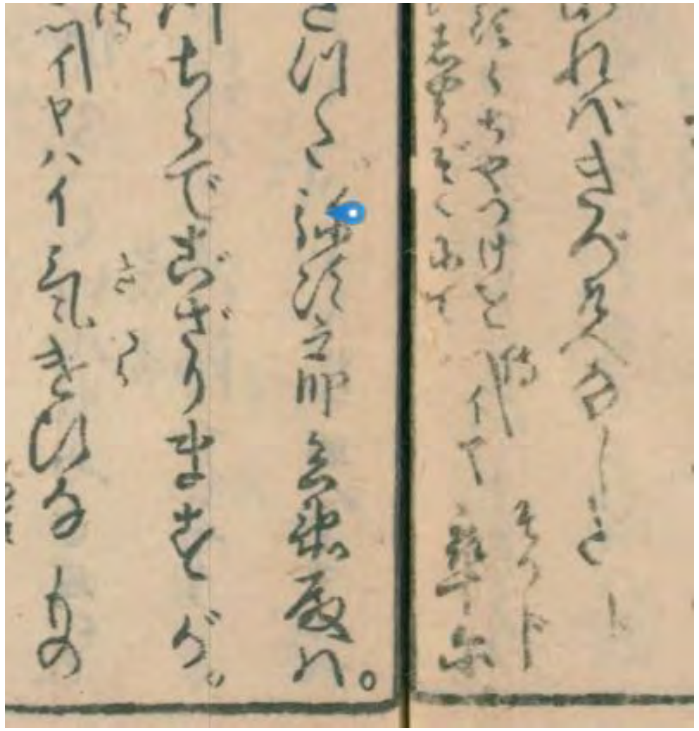
Illustrations in this book

道中膝栗毛 8編続12編 [11]

十返舎一九 作・画椀屋喜兵衛[ほか] 1800

弥次郎兵衛丑年山田の町の□□□□□安土山子見ゆるぬけ□参りニともスルカ府中愚者一徳十一月十一日太夫用主所万文用立所喜多八は。かの上方ものと打つれ。
一人の侍弥次郎兵衛にちかづき「モシあなたがさぬおしの手代いづれ人。おこしでござりますな〽しれた事。太神宮さまへまいりやす〽イヤ太夫ハどれへ〽太夫は竹本義太夫殿さ〽ハア義太夫と申すは。
アノ江戸はどこじやいな〽ゑどは神田の八丁堀とちめんやの弥次郎兵衛さまといつちやァ。ちとひねくつたぬさまだア〽そのおゑどの神田八丁ぼり。とちめんやの弥次郎兵衛どのといふひねくつたやつこさまが。

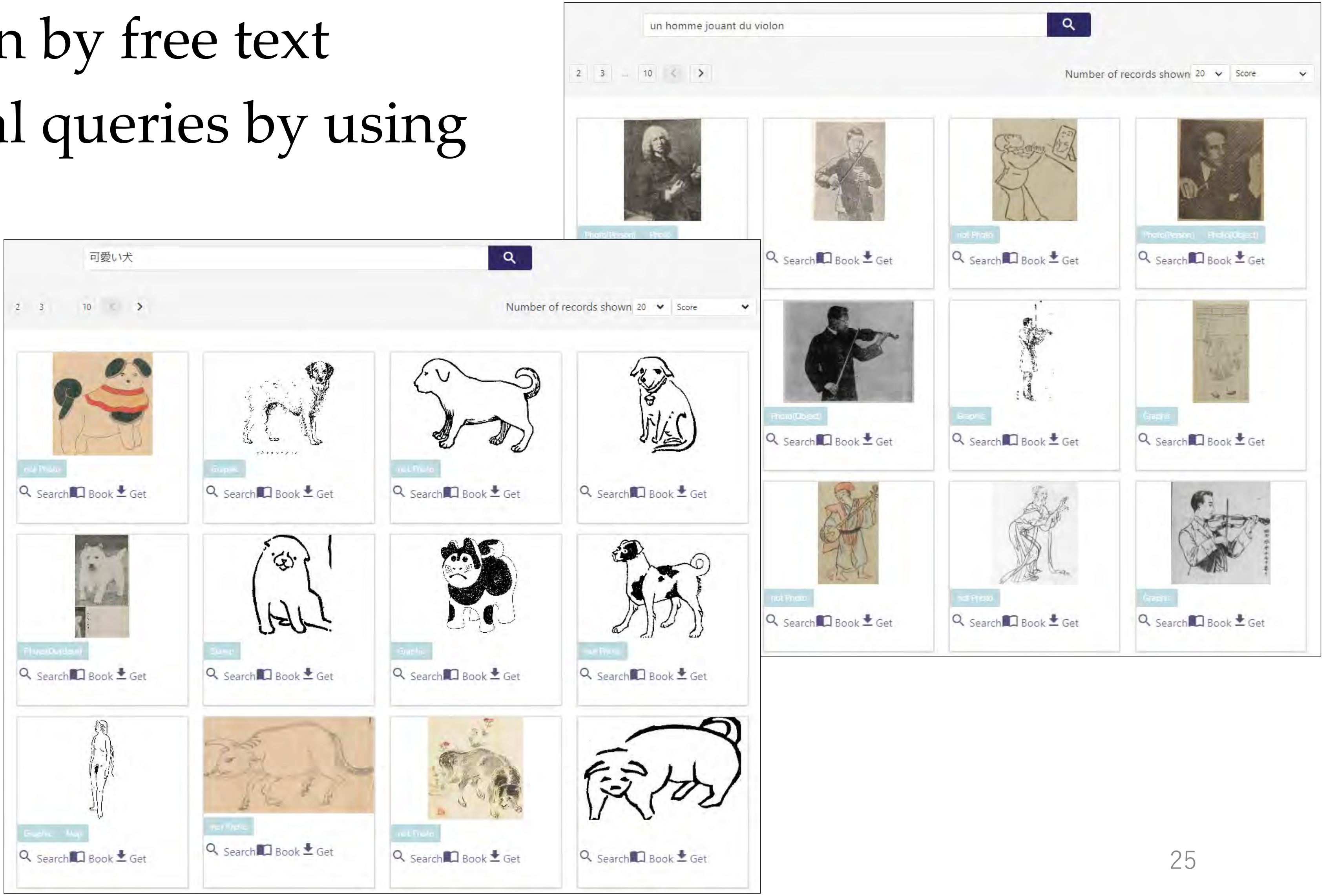
Illustrations in this book



Search for “弥次郎兵衛”

Next Digital Library: “Text-to-Image” Search

- Image search function by free text
- Supports multilingual queries by using machine translation


un homme jouant du violon

Number of records shown 20 Score

可愛い犬

Number of records shown 20 Score

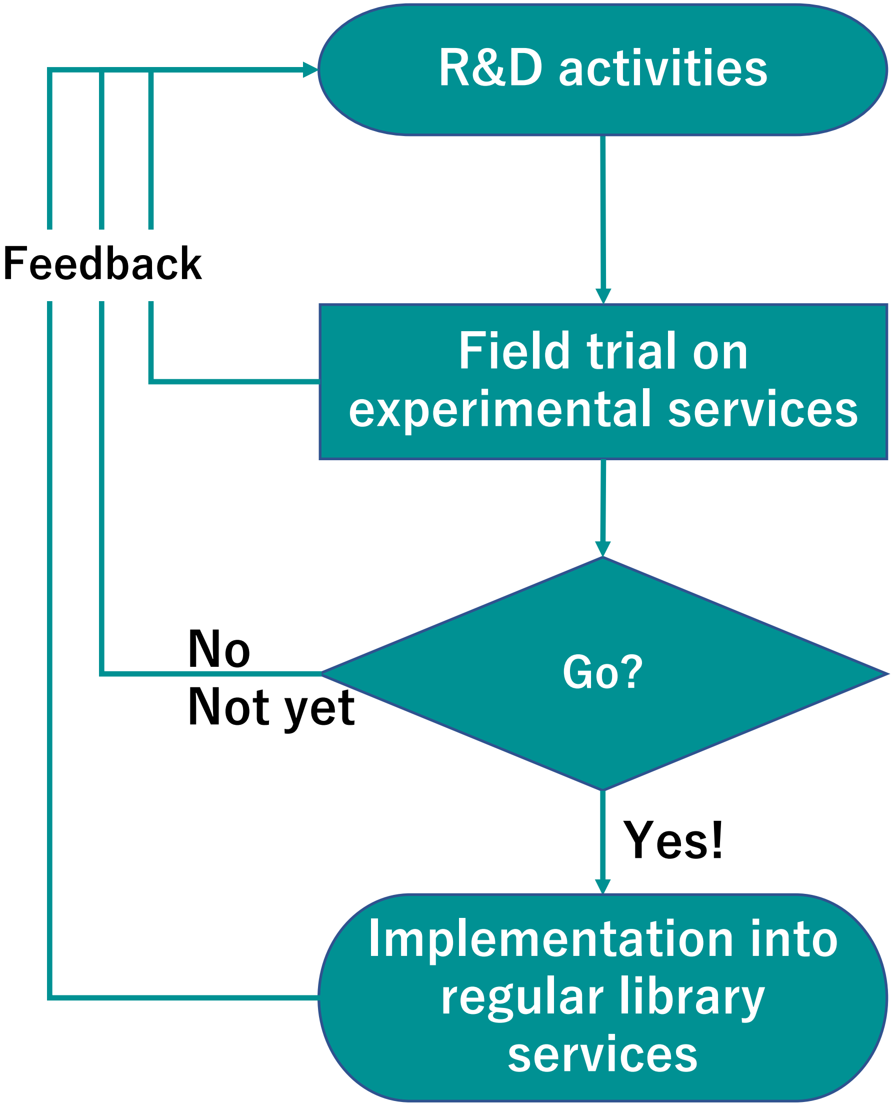
Examples of:

“可愛い犬” (cute dog(s))

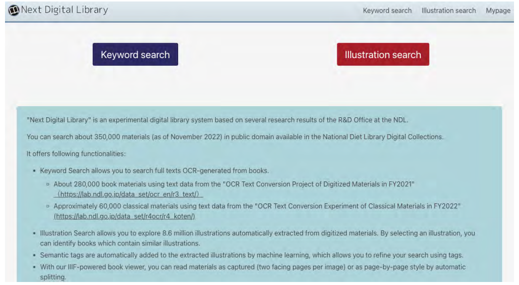
and

“un homme jouant du violon” (a man playing the violin)

Ecosystem Towards New Library Services


R&D activities
Field trial on experimental services
Go?
No
Not yet
Yes!
Implementation into regular library services
Feedback

Next Digital Library



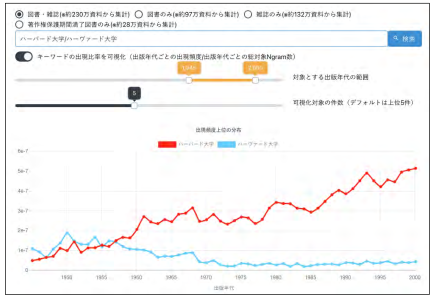
"Next Digital Library" is an experimental digital library system based on several research results of the R&D Office at the NDL.

You can search about 350,000 materials (as of November 2022) in public domain available in the National Diet Library Digital Collections.

It offers following functionalities:

- Keyword Search allows you to search full texts OCR-generated from books.
  - About 280,000 book materials using text data from the "OCR Text Conversion Project of Digitized Materials in FY2021" （https://lab.ndl.go.jp/data_set/ocr_en/r3_text/）
  - Approximately 60,000 classical materials using text data from the "OCR Text Conversion Experiment of Classical Materials in FY2022" (https://lab.ndl.go.jp/data_set/r4ocr/r4_koten/)
- Illustration Search allows you to explore 8.6 million illustrations automatically extracted from digitized materials. By selecting an illustration, you can identify books which contain similar illustrations.
- Semantic tags are automatically added to the extracted illustrations by machine learning, which allows you to refine your search using tags.
- With our IIIF-powered book viewer, you can read materials as captured (two facing pages per image) or as page-by-page style by automatic splitting.

NDL Ngram Viewer



● 図書・雑誌(※約230万資料から集計) ○ 図書のみ(※約97万資料から集計) ○ 雑誌のみ(※約132万資料から集計)
○ 著作権保護期間満了図書のみ(※約28万資料から集計)

ハーバード大学/ハーヴァード大学

キーワードの出現比率を可視化（出版年代ごとの出現頻度/出版年代ごとの総対象Ngram数）

対象とする出版年代の範囲

可視化対象の件数（デフォルトは上位5件）

…

NDL Digital Collections

☑ Available without login ☑ Available with Digitized Contents Transmission Service ☑ Available only at the NDL

Title, Creator…

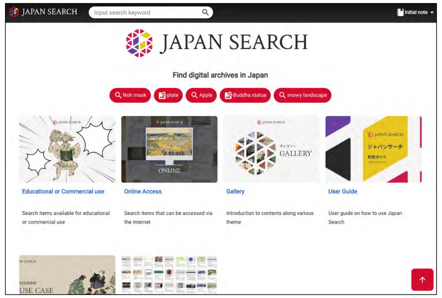
For example... 囲碁 雪まつり 吾輩は猫である シナリオ カット集

News
画像検索不具合のお知らせ（2023-02-15）
現在、画像ファイルや画像ファイルのURL

News
図書約10万点を「国立国会図書館デジタルコレクション」に追加しました（2023-02-03）

News
個人向けデジタル化資料送信サービス（個人送信）に印刷機能が加わりました（2023-01-18）

Japan Search



Find digital archives in Japan

Educational or Commercial use
Search items available for educational or commercial use

Online Access
Search items that can be accessed via the Internet

Gallery
Introduction to contents along various theme

User Guide
User guide on how to use Japan Search

…