

# 東京大学史料編纂所出版物を用いた NDLOCRの応用事例紹介

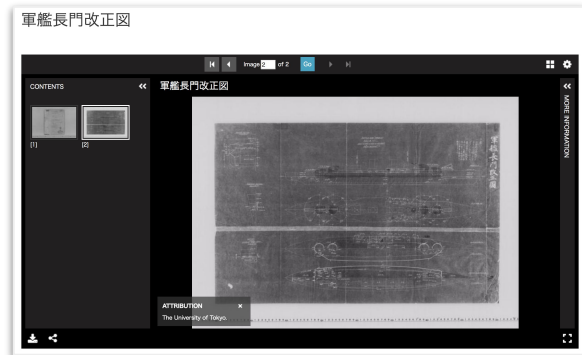
中村覚(東京大学史料編纂所)

# 自己紹介: 中村 覚(なかむら さとる)

- 2012.3: 東京大学 工学部 システム創成学科 卒業
- 2017.3: 東京大学大学院 新領域創成科学研究科 博士課程修了
- 2017.4 - 現在: 東京大学情報基盤センター 助教
  - 2019.4 - 現在: 国立国会図書館 非常勤調査員
- 2020.7 - 現在: 東京大学史料編纂所 助教
- 専門分野: 情報学、人文情報学(Linked Data、デジタルアーカイブ)

# 平賀譲デジタルアーカイブ

- 平賀譲 ひらが・ゆずる(1878-1943)
  - 海軍造船中将
  - 第13代東京帝国大学総長(1938-1943)
- 彼が遺した艦艇計画・建造関係の技術資料を中心とする史料44,000点をデジタル化し公開



# 発表内容

- 東京大学史料編纂所出版物への応用
- 幕末維新史料・横断検索システム

## 参考

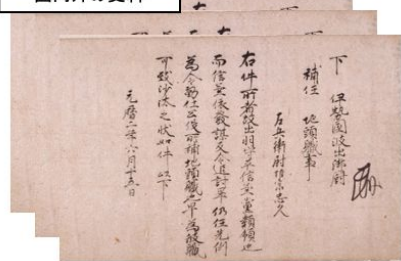
- 中村覚, 劉冠偉, 山田太造. NDLOCRを用いた東京大学史料編纂所史料集版面画像に対する検索システムの開発. 研究報告人文科学とコンピュータ(CH), Vol.2022-CH-130, No.5, pp.1-8, 2022年8月
- 中村覚. OCRテキストを用いた幕末維新史料・横断検索システムの開発. 明治維新史学会例会(2022年10月)

# 東京大学史料編纂 所出版物への応用

# 東京大学史料編纂所における研究活動

## 史料探訪

国内外の史料



## 収集・調査

メタデータ

- ・文書名：源頼朝下文
- ・所蔵先：東京大学史料編纂所
- ・和暦：元暦2年6月15日
- ・・・・

デジタル画像

- ・JPEG 9184x7239 8-bit sRGB

## 編纂

翻刻

再編成

索引

網文

下 伊勢國波出御厨  
補任 地頭職事  
右件所者故出羽守平信兼黨類領也  
而信兼依發謀令追討平仍任先例  
為令勤仕公役所補地頭職也早為發願  
可致沙汰之狀如件以下

オブジェクト識別・同定

- ・史料
- ・人名
- ・地名
- ・・・・

## 出版

平信兼  
所發願  
類領

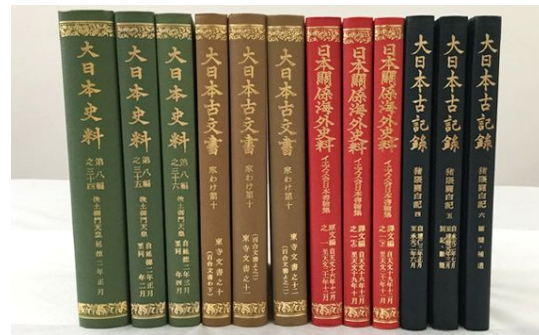
下 伊勢國波出御厨  
補任 地頭職事  
(異書)  
左兵衛尉惟宗忠久

(頭題)  
(花押)

例：『大日本古文書』

- ・史料集
- ・DB公開

右件所者故出羽守平信兼黨類領也而信兼依發謀反令追討畢仍任先例為令勤仕公役所補地頭職也早為彼職可致沙汰之狀如件以下



約200件/年の史料探訪  
共共拠点との連携

# 史料集版面ギャラリー

- **課題:** 史料集版面の画像が公開されているのみ
- OCRによるテキスト化、およびその活用を行いたい
  - 史料集の本文については、編纂所が公開する「大日本史料総合データベース」でも一部利用可能であるが、出版済みの史料集すべてが対象とはなっていない
  - 史料集のシリーズを横断した検索や分析を行いたい

## 史料集版面ギャラリー



東京大学史料編纂所により編纂・出版した史料集の版面画像ギャラリーです。編纂・出版の詳細は[こちら](#)でご確認ください。

クリエイティブ・コモンズ・ライセンスの「CC BY-NC-SA」([クリエイティブ・コモンズ 表示・非営利・継承 4.0 国際ライセンス](#)) 相当の条件で提供しています。詳細は[こちら](#)でご確認ください。

### 『大日本史料』 [1] ⇄

#### 1編 [1] ⇄

- |   |   |
|---|---|
| ▪ <a href="#">1 仁和 3年 8月～寛平3年12月</a> ⇄  | ▪ <a href="#">2 寛平 4年正月～延喜元年10月</a> ⇄   |
| ▪ <a href="#">3 延喜元年11月～8年雑載</a> ⇄      | ▪ <a href="#">4 延喜 9年正月～18年6月</a> ⇄     |
| ▪ <a href="#">5 延喜18年 7月～延長5年10月</a> ⇄  | ▪ <a href="#">6 延長 5年11月～承平5年雑載</a> ⇄   |
| ▪ <a href="#">7 承平 6年正月～天慶4年8月</a> ⇄    | ▪ <a href="#">8 天慶 4年 9月～天曆元年5月</a> ⇄   |
| ▪ <a href="#">9 天曆元年 6月～7年7月</a> ⇄      | ▪ <a href="#">10 天曆 7年 8月～応和元年11月</a> ⇄ |
| ▪ <a href="#">11 応和元年12月～康保4年5月</a> ⇄   | ▪ <a href="#">12 康保 4年 5月～安和2年8月</a> ⇄  |
| ▪ <a href="#">13 安和 2年 8月～天禄3年6月</a> ⇄  | ▪ <a href="#">14 天禄 3年 7月～天延2年3月</a> ⇄  |
| ▪ <a href="#">15 天延 2年 4月～貞元元年6月</a> ⇄  | ▪ <a href="#">16 貞元元年 7月～2年是歳</a> ⇄     |
| ▪ <a href="#">17 貞元 2年雑載～天元3年6月</a> ⇄   | ▪ <a href="#">18 天元 3年 7月～4年10月</a> ⇄   |
| ▪ <a href="#">19 天元 4年11月～5年12月</a> ⇄   | ▪ <a href="#">20 天元 5年閏12月～永観2年2月</a> ⇄ |
| ▪ <a href="#">21 永観 2年 3月～雑載</a> ⇄      | ▪ <a href="#">22 寛和元年正月～3月</a> ⇄        |
| ▪ <a href="#">23 寛和元年 4月～12月</a> ⇄      | ▪ <a href="#">24 寛和元年是歳～2年6月</a> ⇄      |
| ▪ <a href="#">別1 (補遺) 仁和 3年～延喜18年</a> ⇄ | ▪ <a href="#">別2 (補遺) 仁和 3年～延喜21年</a> ⇄ |

# 比較例

NDL OCR

Google Cloud Vision

頭注を正しく認識できている

割書を正しく認識できている

大日本史料 第一編之

宇多天皇

仁和三年丁未

八月 壬寅

二十六日 御踐祚アラセラル

日本紀略 宇多天皇

八月廿六日丁卯立爲皇太子即日受天祚年廿

西宮記 宇多天皇

二刻先帝登殿御侍藤原源光及御劍請太政大臣中其由持進於歷

景嚴少納言左近少將等造資鑑印公卿候常寧殿南廳典侍給子鑑印奉御座

邊郎公卿侍務固關事詔太政大臣率萬事云々

扶桑略記 宇多天皇

八月廿六日丁卯踐祚于時年廿

仁和三年八月二十六日

頭注を正しく認識できている

割書を正しく認識できている

大日本史料 第一編之

宇多天皇

仁和三年丁未

八月 壬寅

二十六日 御踐祚アラセラル

日本紀略 宇多天皇

八月廿六日丁卯立爲皇太子即日受天祚年廿

西宮記 宇多天皇

二刻先帝登殿御侍藤原源光及御劍請太政大臣中其由持進於歷

景嚴少納言左近少將等造資鑑印公卿候常寧殿南廳典侍給子鑑印奉御座

邊郎公卿侍務固關事詔太政大臣率萬事云々

扶桑略記 宇多天皇

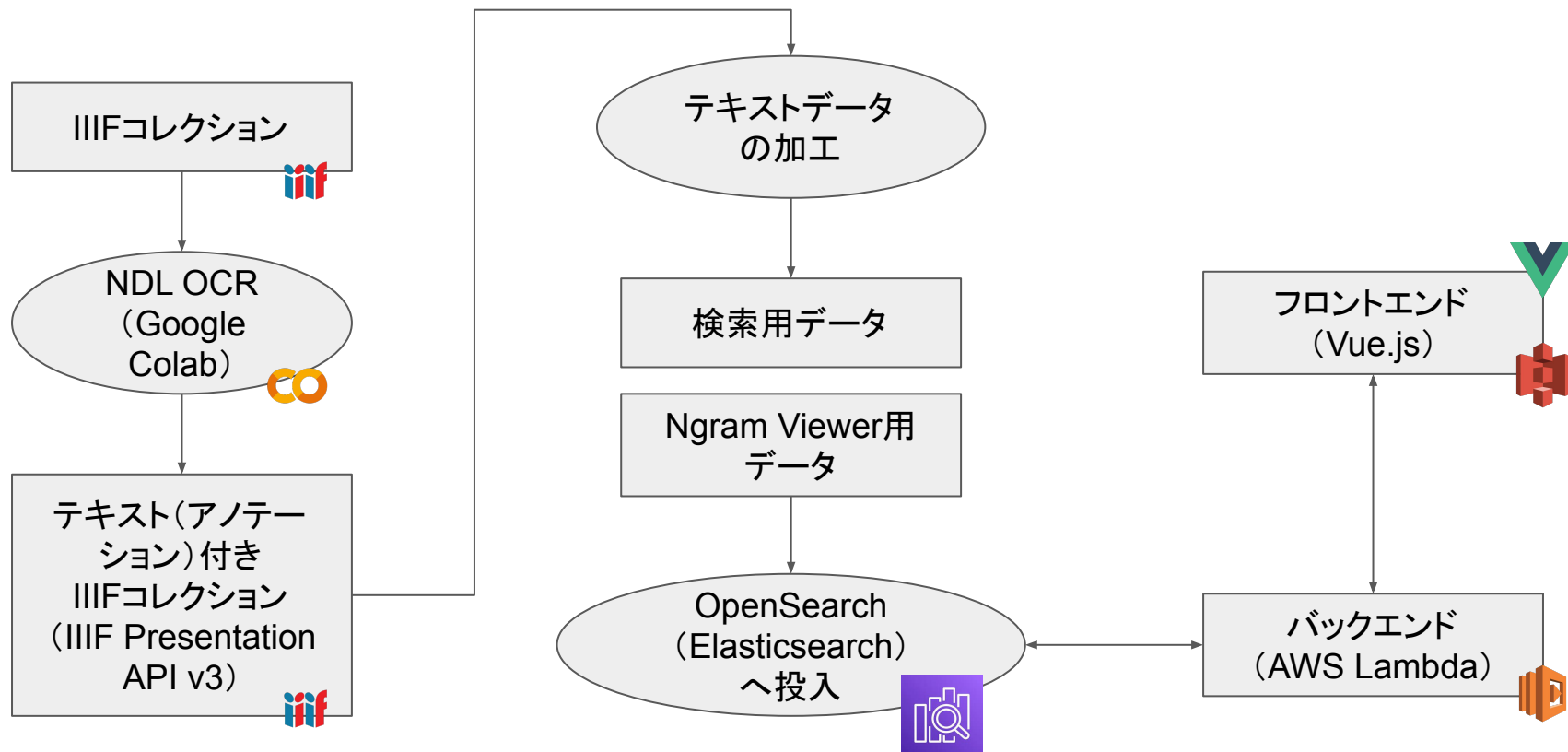
八月廿六日丁卯踐祚于時年廿

仁和三年八月二十六日



# データ構築

# 構築の流れ



# NDLOCRを用いたOCR処理

- **入力:** IIIF マニフェストファイルを入力ファイルとしてOCR処理を適用可能な環境を Google Colaboratory を用いて作成
  - 作成したノートブックは一般に公開 <https://zenn.dev/nakamura196/articles/b6712981af3384>
- **出力:** 入力データである IIIF マニフェストファイルと出力データである XML ファイルを組み合わせることで, OCR テキストをアノテーションの形式で格納する IIIF マニフェストファイルを作成
  - IIIF Presentation API 3.0 を採用
  - IIIF Content Search API 1.0 に準拠したマニフェストファイルおよび API を開発
  - OCR 処理結果を IIIF コレクションの形でまとめる => データセットとしての共有を支援

# NDL OCRの適用結果の確認(IIIFコレクション)

## 内容

- NDL OCRの結果をIIIFのアノテーション(IIIF Presentation API v3)に変換
- さらに、IIIF Content Search APIを作成し、検索機能を追加

上記はIIIFの仕様に準拠して構築しているため、IIIF対応のビューアが提供する表示や検索機能を利用できる。

# テキストデータの加工

- (1) OCR結果を用いた検索インデックスの作成
- (2) OCR結果からの試験的な情報抽出
- (3) Ngram Viewer用のデータ作成

# (1) OCR結果を用いた検索インデックスの作成

表 1 検索用のデータ項目例

Table 1 Example of data items for search.

property	説明	例
Title	タイトル	1 明徳3年閏10月～応永2年3月 p.12
Category1	第一階層	『大日本史料』
Category2	第二階層	『大日本史料』 7編
Category3	冊タイトル	『大日本史料』7編 1 明徳3年閏10月～応永2年3月
Manifest	IIIF マニフェストファイル	<a href="https://.../0701/manifest.json">https://.../0701/manifest.json</a>
Canvas	カンバス	<a href="https://.../0701/canvas/p12">https://.../0701/canvas/p12</a>
Element	要素	["図表", "頭注", ...]
頭注	頭注	["三種神器", "回祿ノ疑", ...]

```
<LINE CONF="0.940" HEIGHT="39" STRING="卯" TYPE="割注" WIDTH="39" X="1203" Y="950"/>
<LINE CONF="0.608" HEIGHT="41" STRING="丁" TYPE="割注" WIDTH="37" X="1320" Y="954"/>
<LINE CONF="1.000" HEIGHT="158" STRING="神器ヲ皇" TYPE="頭注" WIDTH="39" X="1123" Y="403"/>
<LINE CONF="1.000" HEIGHT="154" STRING="太子ノ直" TYPE="頭注" WIDTH="36" X="1082" Y="404"/>
<LINE CONF="0.999" HEIGHT="76" STRING="固關" TYPE="頭注" WIDTH="36" X="554" Y="403"/>
<LINE CONF="0.999" HEIGHT="146" STRING="曹二奉ル" TYPE="頭注" WIDTH="36" X="1042" Y="404"/>
<BLOCK CONF="0.964" HEIGHT="389" STRING="仁和三年八月二十六日" TYPE="柱" WIDTH="42" X="233" Y="403"/>
<BLOCK CONF="0.996" HEIGHT="82" TYPE="ルビ" WIDTH="30" X="812" Y="864"/>
<BLOCK CONF="0.991" HEIGHT="105" TYPE="ルビ" WIDTH="26" X="596" Y="1741"/>
```

図 2 NDLOCR の出力結果の例

Figure 2 Example of NDLOCR output results.

## (2) OCR 結果からの試験的な情報抽出

### 年月日の抽出

- アノテーションタイプ「柱」の文字列を対象として、正規表現による年月日の抽出
- 抽出した年月日の文字列を HuTime Web APIを用いて西暦表記に変換
  - この処理において、抽出された年月日として不適切な文字列は除外される(例: == 三年三月)
  - 「南朝正平二十三年 北朝應安元年雜載」等の柱書きに対しては現在未対応

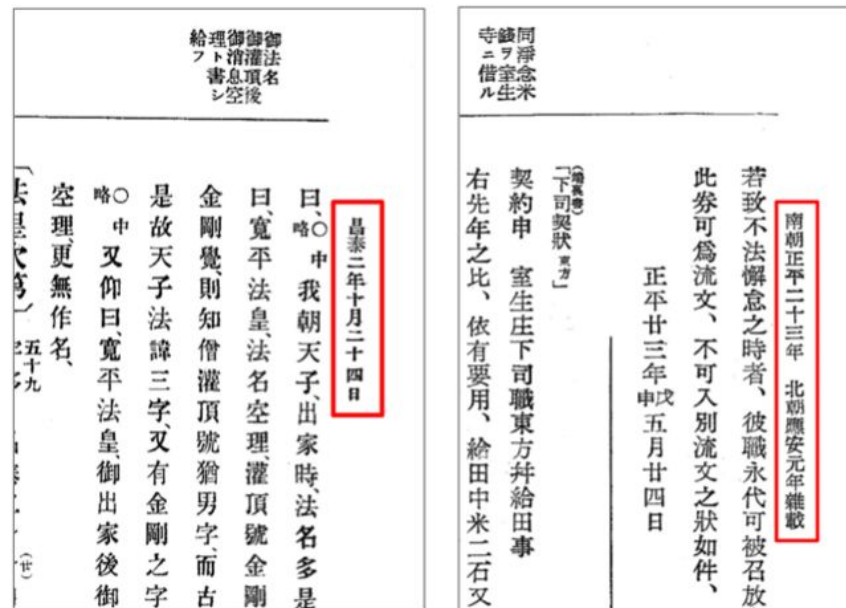


図 3 NDLOCR の TYPE 属性が「柱」の例

Figure 3 Example of NDLOCR with TYPE attribute of “柱”

## (2) OCR 結果からの試験的な情報抽出

### 地名などの固有表現の抽出

固有表現の抽出に関する初歩的な取り組みとして、事前に準備した辞書を用い、辞書に含まれる文字列を抽出する

- 地名
  - 歴史地名データを利用して、クラスが「行政地名/国」である地名(例:大和国, 山城国)





### (3) Ngram Viewer 用のデータ作成

- OCR 結果から 1gram から 5gram までの文字列を作成し、検索インデックスに登録
- 各 Ngram 文字列に対して、上述した処理によって抽出した年毎の出現頻度をプロパティとして登録
  - これにより、年毎の出現頻度の変化の可視化を行う

# 開発したシステム

# 提供機能

- テキスト検索
- Ngram Viewer
- Googleデータポータルを用いた統計分析例
- IIFコレクションを用いたOCR結果の確認ページ

# 検索画面

地震

詳細検索

検索オプション①

☒ 通常検索

☐ 正規表現

☐ 曖昧フレーズ検索1

☐ 曖昧フレーズ検索2

検索オプション②

☒ AND検索

☐ OR検索

年: 1847


クリア

検索結果: 72

1 / 4

20 件

適合度 降順




### 4 弘化4年2月～同年3月 p.689

第一階層:『大日本維新史料 編年之部』 第二階層:『大日本維新史料 編年之部』 1編 史料名:『大日本維新史料 編年之部』 1編 4 弘化4年2月～同年3月

○信越地震記事ニモ略々同一内容ノ記事アリ。

在町破損所御坐候、遠在し儀は、右し通御座候、此段御届申上候、以上、私在所信濃國松本、去ル廿四日夜四ツ時頃地震強、翌廿五日に爲差儀は無之、間遠二相、(信越地震記事)、未相分兼申候、先此段御届申上候、以上、以上松本藩主戸田光則届書)、三月廿七日在所日附、松平丹波守、男拾壹人、一死人六十七人内七人焼死、松本領分地震異變取調三月廿七日在所日附松平丹波守、○帝國圖書館所蔵本、○中村■、六所蔵本、胡路■所載、信州松本地、方地震



### 4 弘化4年2月～同年3月 p.709

第一階層:『大日本維新史料 編年之部』 第二階層:『大日本維新史料 編年之部』 1編 史料名:『大日本維新史料 編年之部』 1編 4 弘化4年2月～同年3月

二有之、夫と不絶震動い■し、折々地震發シ、翌廿五日卯中刻、漸々相、上ケ可申を難斗、同廿六日に至り候なも并人馬死亡等も有之趣、其上篠井川筋兩岸拔崩、水中埋押、水吐兼田方水溢追々畑方に押、弘化四末年三月大地震し儀二付御届書、出雲守領分信濃國高井郡・水内郡み内村々、去月廿四日亥刻頃大地震二な、人家震ヒ潰、三月晦日地震し儀二付御届、堀出雲守家來、四月四日、奥田爲太郎、四月四日堀出雲守家來、○胡路■、隔■所載、讓隔■、震島御届書對、○胡路、○信濃國大地、被害アリ、信州中野地、方地震

# 版面画像の表示

閉じ	7 承平 6年正月
第一階層	『大日本史料』
第二階層	『大日本史料』
史料名	『大日本史料』
〔貞信公記抄〕五月六日大地震、後聊振鳴、于今不止、	
七日、地震、	
八日、地震、	
十日、地震、	
廿一日、大地震、後時々振鳴、不休于今、	
廿二日、戊辰、地震、	
廿三日、己巳、地震、	
廿四日、地震、	
廿五日、辛未、地震數々、	
廿六日、夜地震、	
廿七日、地大震、	
廿九日、地震、	
〔日本紀略〕 <small>宋雀院</small> 五月廿六日、壬申 <small>中略</small> 地震、	
○七日以後ノ地震便宜合敘ス、臨時御讀經ヲ修シ ルコト、四月二十六日ノ條ニ、震災ヲ祈禱センガ爲	
天慶元年五月六日	

# 「図版」を含むページの検索

検索結果: 3,620

2 / 181 < >

20 件 適合度 降順

3 建久元年1月～3年2月 p.854

4 建久3年3月～6年8月 p.110

27 保安2年5月～保安2年11月 p.248

27 保安2年5月～保安2年11月 p.362

28 保安2年11月～保安2年補遺 p.267

2 文治3年9月～建久元年1月 p.556

6 寛弘4年12月～8年6月 p.602

6 寛弘4年12月～8年6月 p.682

6 寛弘4年12月～8年6月 p.685

4 長保3年9月～5年12月 p.15

21 元永元年12月～元永2年2月 p.35

21 元永元年12月～元永2年2月 p.166

# Ngram Viewer

- 複数のキーワードによる検索
- 正規表現による検索
- 表示対象とする年代の範囲の絞り込み



# 統計分析例

- 史料集の特徴を定量的に算出
  - 1アイテムまたはページあたりの文字数
  - 例1:「大日本近世史料」は刊行頻度が高いため、1アイテムあたりの文字数が少ない
  - 例2:「大日本古文書」は図表が多いため、1アイテムあたりの文字数が少ない

東京大学史料編纂所・史料集版面の統計分析例

総文字数	史料数	総ページ数	1史料あたりの平均ページ数	1アイテムあたりの平均文字数
2億	914	49万	495	22万

ID	ラベル	アイテム数	頁数	文字数	1アイテムあたりの頁数	1アイテムあたりの文字数
02	『大日本史料』	407	259664	103744824	637	254901
04	『大日本維新史料 緒年之部』	19	15041	5725644	791	301349
05	『大日本古文書』	226	112686	33088067	498	146407
07	『大日本近世史料』	138	47001	16804756	340	121773
08	『大日本維新史料 類纂之部』	30	10622	4791636	354	159721
11	『復古記』	16	12470	7521868	779	470116
12	『明治史要』	2	926	388404	463	194202
15	『維新史』	7	4442	2246143	634	320877
29	『日本関係海外史料目録』	14	3867	2472332	276	176595
52	『明治維新史料選集』	2	781	288811	390	144405
53	『保古飛呂比』	12	5747	271166	478	225930
58	『日本関係海外史料』	41	12545	5611230	305	136859

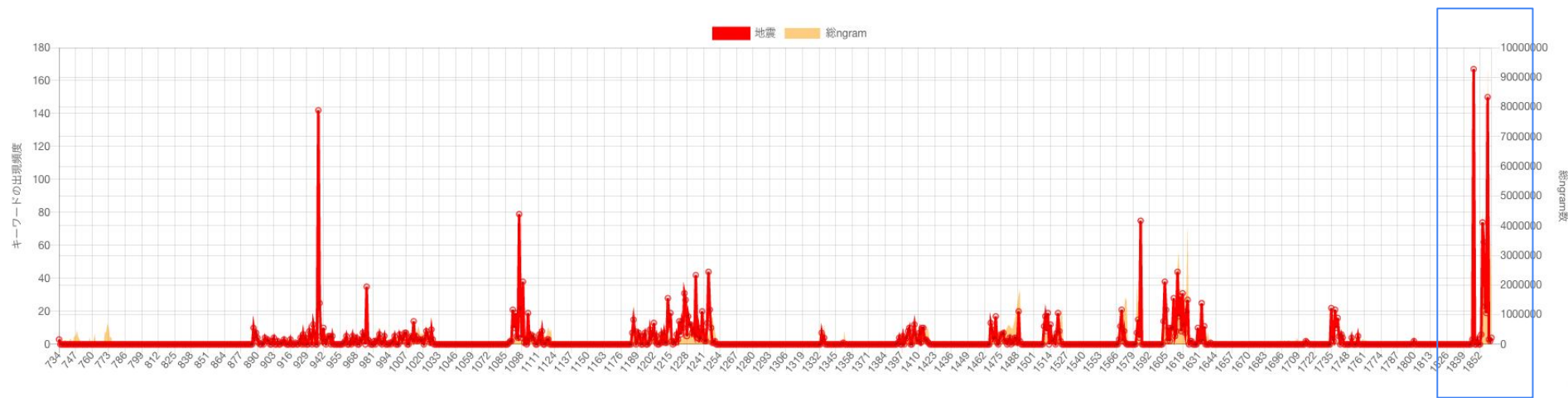


# 利用例

# 1840～1860 までの地震

1847年	弘化4年	善光寺地震	長野県
5月8日	3月24日		
1854年	安政元年	伊賀上野の地震	三重県伊賀市北部
7月9日	6月15日		
1854年	安政元年	安政東海地震	東海
12月23日	11月4日		
1854年	安政元年	安政南海地震	南海
12月24日	11月5日		
1855年	安政2年	安政江戸地震	関東地方南部
11月11日	12月2日		
1858年	安政5年	飛越地震	越中・飛騨国境（現在の富山・岐阜県境）
4月9日	2月26日		

# Ngram Viewerで「地震」を検索



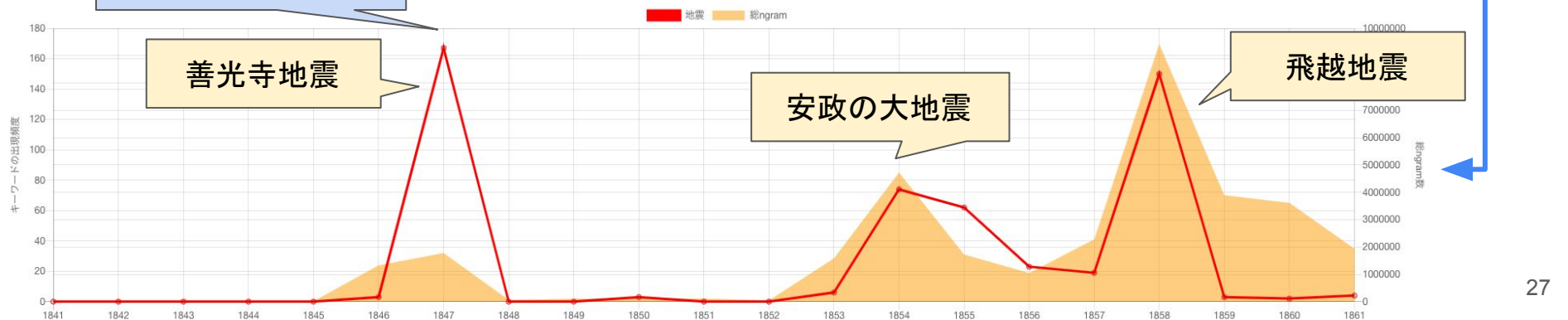
ポイントをクリック

善光寺地震

安政の大地震

飛越地震

拡大表示



地震

## 地震

検索オプション① ☒ 通常検索 ☐ 正規表現 ☐ 曖昧フレーズ検索1 ☐ 曖昧フレーズ検索2

検索オプション② ☒ AND検索 ☐ OR検索

年: 1847

クリア

1847年

検索結果: 72

史料集版面ページ

### 4 弘化4年2月～同年3月 p.689

第一階層: 『大日本維新史料 編年之部』 第二階層: 『大日本維新史料 編年之部』 1編 史料名: 『大日本維新史料 編年之部』 1編

○信越地震記事ニモ略々同一内容ノ記事アリ。

在町破損所御坐候、遠在儀は、右し通御座候、此段御届申上候、以上、私在所信濃國松本、去ル  
、翌廿五日と爲差儀は無之、間遠ニ相、(信越地震記事)、未相分兼申候、先此段御届申上候、以上、  
松本藩主戸田光則届書)、三月廿七日在所日附、松平丹波守、男拾壹人、一死人六十七人内七人焼死、松  
三月廿七日在所日附松平丹波守、○帝國圖書館所藏本、○中村■、六所藏本、胡路■所載、信州松本地

### 4 弘化4年2月～同年3月 p.709

第一階層: 『大日本維新史料 編年之部』 第二階層: 『大日本維新史料 編年之部』 1編 史料名: 『大日本維新史料 編年之部』 1編 4 弘化4年2月～同年3月

ニ有之、夫ち不絶震動い■し、折る地震發シ、翌廿五日卯中刻、漸る相、上ケ可申を難斗、同廿六日に至り候なも  
并人馬死亡等も有之趣、其上篠井川筋兩岸拔崩、水中埋押、水吐兼田方水溢追る畑方に押、弘化四末年三月大地震  
し儀二付御届書、出雲守領分信濃國高井郡・水内郡み内村る、去月廿四日亥刻頃と大地震二な、人家震ヒ潰、  
三月晦日地震し儀二付御届、堀出雲守家來、四月四日、奥田爲太郎、四月四日堀出雲守家來、○胡路  
■、■所載、讓■、震島御届書對、○胡路、○信濃國大地、被害アリ、信州中野地、方地震

ページ内ハイライト

一死人六十七人内七人焼死

松本領分地震異變取調覺

〔信越地震記事〕○中村■  
六所藏本

○信越地震記事ニモ略々同一内容ノ記事アリ。

三月廿七日 在所日附

私在所信濃國松本、去ル廿四日夜四ツ時頃  
成候得共、今以相止不申候城內無別條、  
未相分兼申候、先此段御届申上候、以上、  
伊、委細く後へ

書損松  
本

信州松本地  
方大地震ア  
リ

同四月三日、御用番戸田山崎守候へ差出之

# 実験的な機能: 地名の抽出

東京大学史料編纂所・史料集版面検索

- ☐ 弘化四年二月二十四日 5
- ☐ 弘化四年四月二十八日 5
- ☐ 弘化四年五月三日 2
- ☐ 弘化四年五月二十三日 1
- ☐ 弘化四年五月十七日 1
- ☐ 弘化四年八月十八日 1
- ☐ 弘化四年六月十四日 1
- ☐ 弘化四年四月二十一日 1
- ☐ 弘化四年四月二十三日 1
- ☐ 弘化四年四月二十四日 1

すべて表示 ▶

地名 4

- ☐ 信濃国 22
- ☐ 越後国 8
- ☐ 相模国 1
- ☐ 肥後国 1

すべて表示 ▶

寺 0

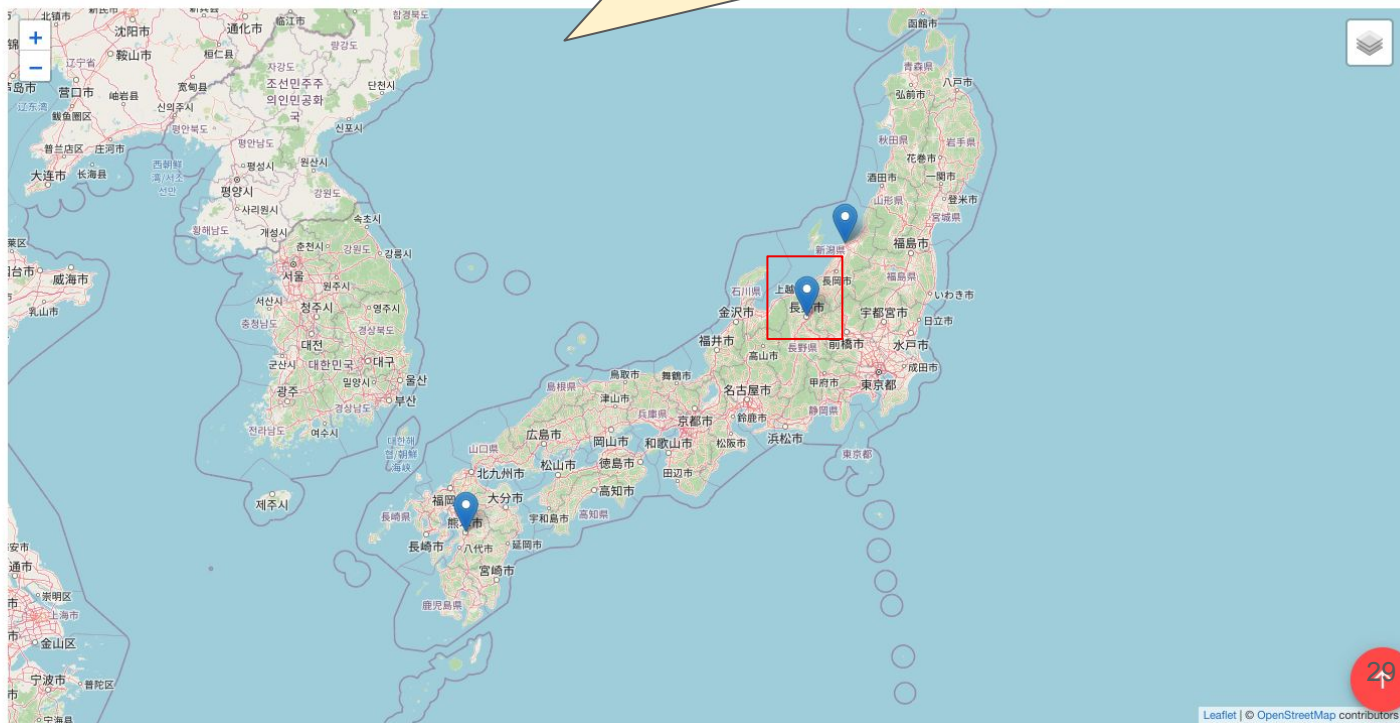
すべて表示 ▶

Role 0

年: 1847 クリア

検索結果: 72

「地震」が含まれる「1847」年の史料



# 幕末維新史料・横 断検索システム

# 幕末維新史料・横断検索システム

東京大学史料編纂所が公開する『大日本維新史料 編年之部』『大日本維新史料 類纂之部』『大日本古文書 幕末外国関係文書』『復古記』『明治史要』『保古飛呂比』『明治維新史料選集』『維新史』などに加えて、『孝明天皇紀』と日本史籍協会叢書のOCRテキストを検索



本システムは、デジタルライジング研究会(科学研究費補助金・基盤研究 (B)「日本近世史料学の再構築 — 基幹史料集の多角的利用環境形成と社会連携を通じて(研究代表者:杉本史子)」(課題番号 22H00692)、2022年6月28日オンライン開催、を契機として開発がスタートしました。

# 検索対象の設定機能

検索対象範囲を利用者がカスタマイズ可能な機能の提供

検索対象史料については、今後ルールなどを設け、適宜追加予定

幕末維新史料・横断検索システム

ホーム > マイサーチ > 検索設定

## 検索設定

### 検索対象

<input type="checkbox"/> 『大日本史料』 (259,664)	<input type="checkbox"/> 『大日本古文書』 (112,686)	<input checked="" type="checkbox"/> 日本史籍協会 (84,530)	<input type="checkbox"/> 『大日本近世史料』 (47,001)
<input checked="" type="checkbox"/> 『大日本維新史料 編年 之部』 (15,041)	<input type="checkbox"/> 『日本関係海外史料』 (12,545)	<input checked="" type="checkbox"/> 『復古記』 (12,470)	<input checked="" type="checkbox"/> 孝明天皇紀 (11,702)
<input checked="" type="checkbox"/> 『大日本維新史料 類纂 之部』 (10,622)	<input checked="" type="checkbox"/> 『保古飛呂比』 (5,747)	<input checked="" type="checkbox"/> 『維新史』 (4,442)	<input type="checkbox"/> 『日本関係海外史料目 録』 (3,867)
<input checked="" type="checkbox"/> 『明治史要』 (926)	<input type="checkbox"/> 『明治維新史料選集』 (781)		

このサイトについて

機能

リンク

言語

お知らせ

検索

史料集版面ギャラリー

English

更新履歴

カテゴリ

東京大学史料編纂所



# 考察

# 利点

- 版面画像に対するテキスト検索
  - 画像のみが公開され、検索が困難であった史料集版面に対して、テキスト検索が可能になった
  - NDLOCRを使用することにより、従来の OCRソフトウェアおよびサービスでは困難であった割書などに対する高精度なテキスト化を実現
- 史料集に対する横断検索
  - OCR処理結果の誤り、および年月日の抽出処理の精度などには課題が残るものの、編纂所が提供するデータベース、および日本史研究に新たな価値を与えるツールとして機能し得る

# 今後の課題

- OCRテキストの誤り修正
  - NDLOCR特有の誤りに対する集中的な訂正方法の検討
  - クラウドソーシング的な手法を用いた人手による修正
  - 編纂所が既に公開しているデータベースの活用
  - 修正に加えて、テキストの構造化を行うための環境も構築したい
- 既存のシステムとの棲み分け
  - 本システムは、2023年1月時点において、関係者限りの共有としています。また当面は実験的なサービスであり、予告無くサービスの中止、改変を行う可能性があります。
  - 本実験サービスの評価・検証を踏まえて、今後の一般公開などを検討予定です。

# 謝辞

本研究はJSPS科研費22H00692および東京大学史料編纂所「データ駆動型歴史情報研究基盤の構築」による成果の一部です。

ご清聴ありがとうございました。