

日本文化をAIとビッグデータで読み解く ー過去の日本文化を「見える化」するデ ジタル技術



北本 朝展 (KITAMOTO Asanobu)

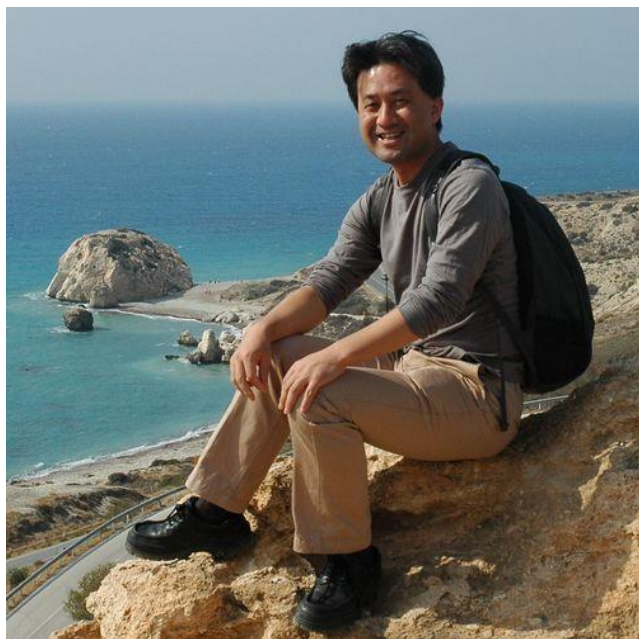
国立情報学研究所

ROIS-DS人文学オープンデータ共同利用センター (CODH)

<https://researchmap.jp/kitamoto/>

kitamoto@nii.ac.jp

自己紹介



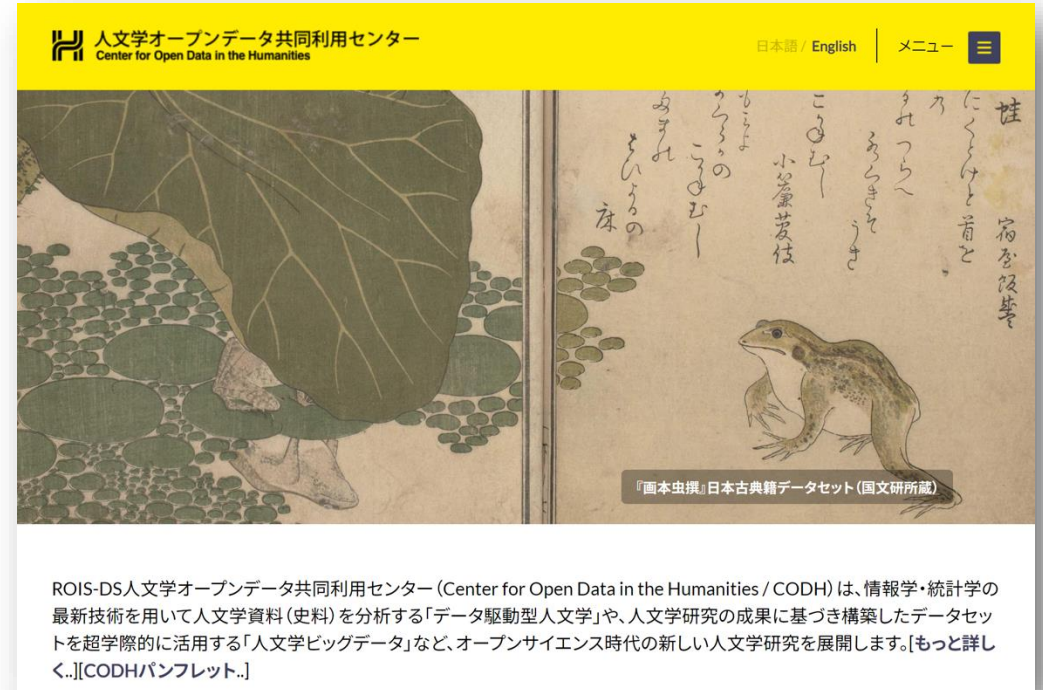
@kitamotoasanobu

<https://researchmap.jp/kitamoto/>

1. 画像処理研究から情報学全般に拡大
2. 「**データ駆動型サイエンス**」のアプローチで、他分野の研究者と協働
 - 気象情報
 - 地球環境情報
 - 地理情報
 - **人文情報**
3. デジタルアーカイブ・オープンサイエンスなどにも関わる

ROIS-DS人文学オープンデータ共同利用センター（CODH）

1. **データ駆動型人文学**：情報学・統計学の最新技術を用いて人文学資料（史料）を分析
2. **人文学ビッグデータ**：人文学研究の成果に基づくデータセットを超学際的に活用
3. **オープンサイエンス**：時代の新しい人文学研究を展開



<http://codh.rois.ac.jp/>

文字・テキストのギャップ



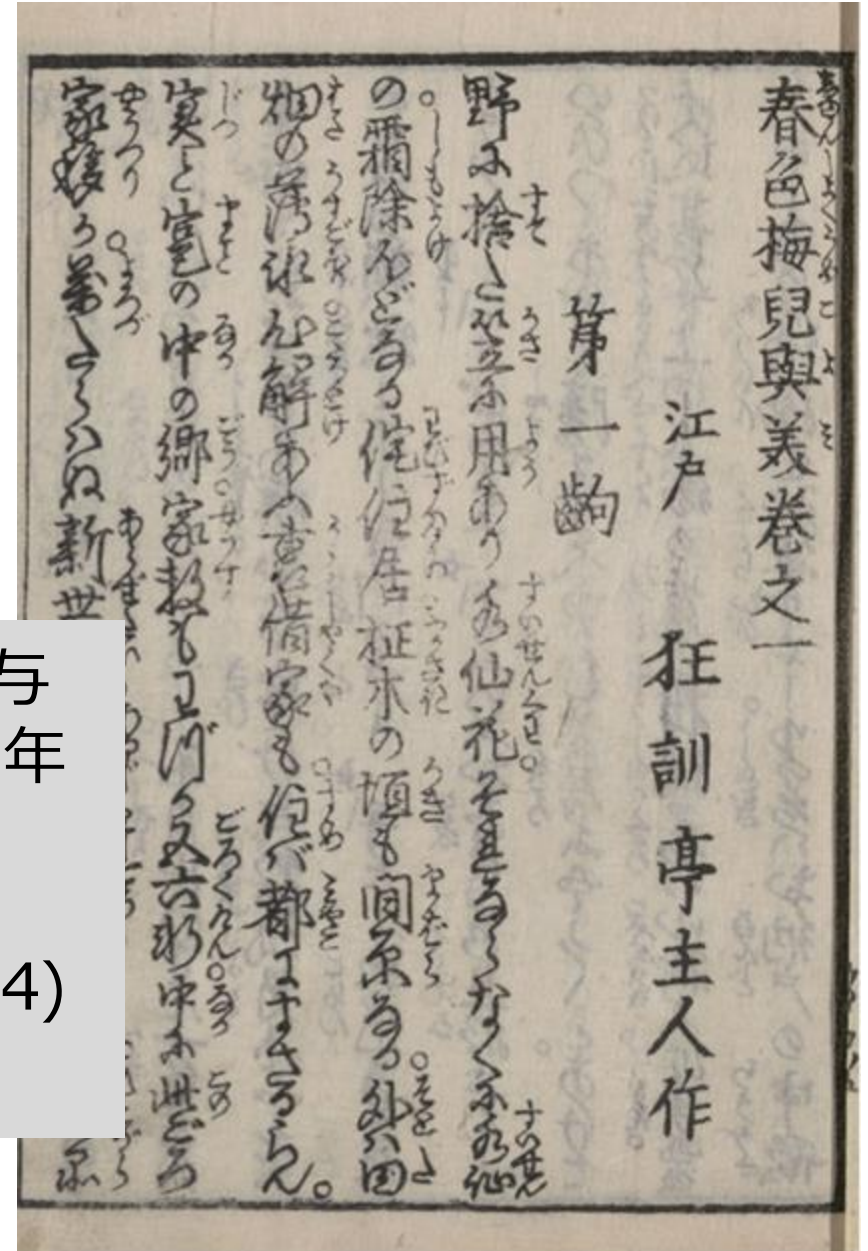
CHAPTER I.

DOWN THE RABBIT-HOLE

ALICE was beginning to get very bored sitting by her sister on the bank, and there was nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is

Alice's adventures in
wonderland (1894
edition)
by Carroll, Lewis,
1832-1898
(archive.org)

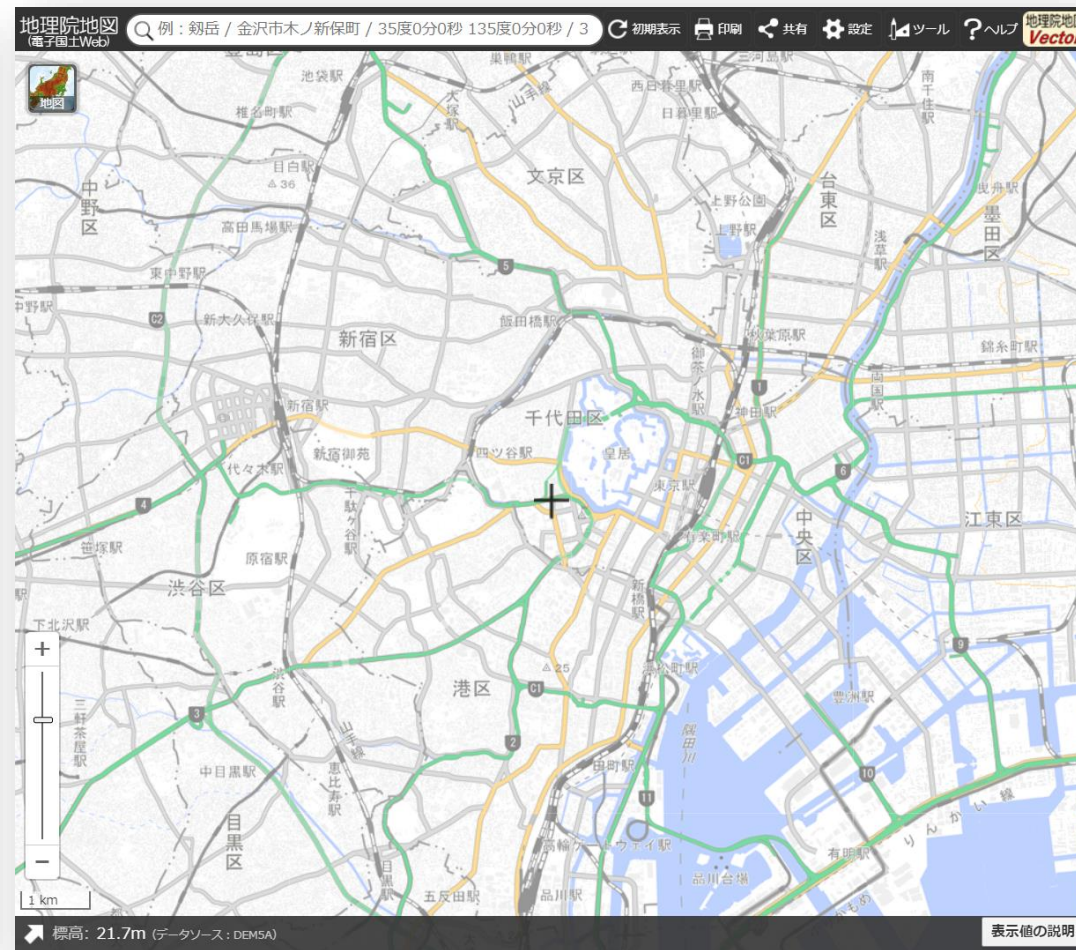
『春色梅児与
美』（1833年
版）
為永春水
(1790-1844)
(国語研蔵)



地理空間のギャップ



れきちず（江戸時代）



地理院地図（現代）

景観のギャップ

メモリーグラフ
<https://mp.ex.nii.ac.jp/>

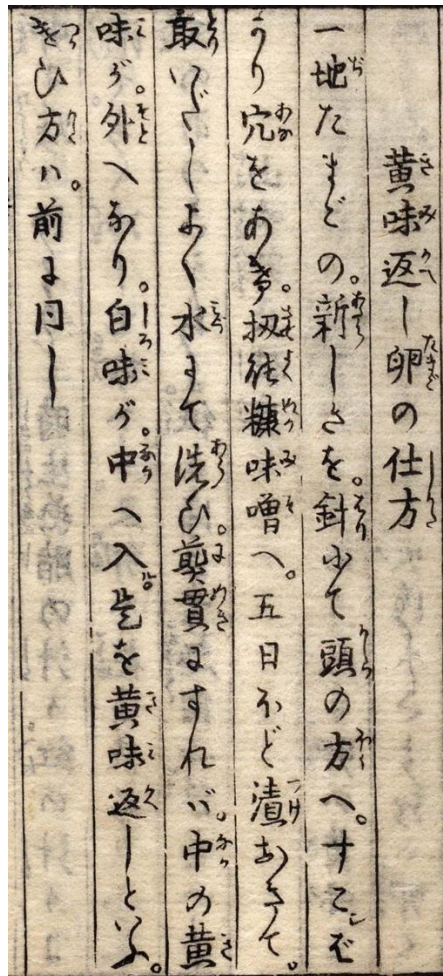


三宅坂より司法省を望む, 写真の中の明治・大正, 国立国会図書館

江戸料理レシピデータセット

<http://codh.rois.ac.jp/edo-cooking/>

『万宝料理秘密箱
卵百珍』, 1795



- | | |
|---|--------------------------------------|
| 1 | 地たまごの 新しきを 針にて頭(かしら)の方へ すこしばかり穴をあけ |
| 2 | 扱能(さてよく)糠味噌へ 五日ほど漬おきて |
| 3 | 取だしよく水にて洗ひ |
| 4 | 煎貫にすれば 中の黄味が 外へなり 白味が 中へ入ル 是を黄味返しといふ |



翻刻：現代文字に変換
翻訳：現代日本語に変換
レシピ化：アクション可能な情報に変換

世界に一つだけの黄身返し卵

考案：三ツ星たまごソムリエ友加里

一度たりとも同じ模様にならないところが、面白い魅力である黄身返し卵。現代の身近な道具で再現してみました。



過去と現在のギャップ

1. 人文学研究では、過去の文化を過去の文脈で理解することが専門性となる
2. 専門外の人には、過去の文脈だけでは理解できないため、大きなギャップが生じる
3. 過去の文化を新しいプラットフォームに入れる = 現在の人々になじみ深い方法で過去の文化にアクセス
4. 過去と現在のギャップを縮めるためのAI（人工知能）やシステムを開発



国文学研究資料館 大型プロジェクト（DDHプロジェクト）

<https://lab.nijl.ac.jp/humanitiesthroughddps/>

**300,000点の日本古典籍
（1868年以前）を、デジ
タル化し、オープンデー
タとして公開**

日本文化のビッグ
データをどのよう
に活用するか？

過去の文化遺産と「くずし字」の問題

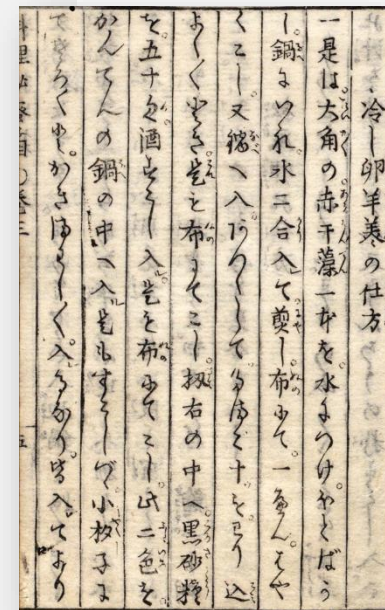


文書
数億点

日本に残された古典
籍や古文書の点数
(冊数) の推定

読者
数千人

くずし字をきちんと
読める人数の推定
(**全人口の0.01%**)



変体仮名

<http://codh.rois.ac.jp/char-shape/hentaigana/>

あ	あ - 安 U+1B002	愛 - 愛 U+1B003	阿 - 阿 U+1B004	惡 - 惡 U+1B005		
い	以 - 以 U+1B006	伊 - 伊 U+1B007	意 - 意 U+1B008	移 - 移 U+1B009		
う	宇 - 宇 U+1B00A	宇 - 宇 U+1B00B	憂 - 憂 U+1B00C	有 - 有 U+1B00D	雲 - 雲 U+1B00E	
え	江 - 江 U+1B001	盈 - 盈 U+1B00F	縁 - 縁 U+1B010	衣 - 衣 U+1B011	衣 - 衣 U+1B012	要 - 要 U+1B013
お	於 - 於 U+1B014	於 - 於 U+1B015	隱 - 隱 U+1B016			
か	佳 - 佳 U+1B017	加 - 加 U+1B018	可 - 可 U+1B019	可 - 可 U+1B01A	嘉 - 嘉 U+1B01B	
	我 - 我 U+1B01C	歟 - 歟 U+1B01D	賀 - 賀 U+1B01E	閑 - 閑 U+1B01F	香 - 香 U+1B020	
	駕 - 駕 U+1B021	家 - 家 U+1B022				
き	喜 - 喜 U+1B023	幾 - 幾 U+1B024	幾 - 幾 U+1B025	支 - 支 U+1B026	木 - 木 U+1B027	
	祈 - 祈 U+1B028	貴 - 貴 U+1B029	起 - 起 U+1B02A	期 - 期 U+1B03B		

1. 現代のひらがなは、
一音一字。「あ」は
一つしかない
2. 現代の仮名以外を
「**変体仮名**」という
3. **1900年の小学校令**に
より、1種類以外は教
えなくなった
4. 現代の日本人はほと
んど読めない！

日本古典籍くずし字データセット

<http://codh.rois.ac.jp/char-shape/>



くずし字データセット

Unicode	Image	X	Y	Width	Height
U+842C	200021853-00002	634	244	127	163
U+5BB6	200021853-00002	645	424	123	156
U+65E5	200021853-00002	665	611	65	87
U+7528	200021853-00002	650	727	97	123
U+60E3	200021853-00002	644	883	121	140
U+83DC	200021853-00002	640	1048	120	164
U+4FCE	200021853-00002	638	1249	136	124
U+4E0D	200021853-00002	468	260	127	108
U+6642	200021853-00002	477	383	124	145
U+73CD	200021853-00002	462	545	151	129
U+5BA2	200021853-00002	466	692	136	141
U+5373	200021853-00002	472	851	124	124
U+5E2D	200021853-00002	465	985	132	145
U+5E96	200021853-00002	469	1149	133	131
U+4E01	200021853-00002	480	1288	121	100
U+5408	200021853-00002	533	1553	179	127

CSV形式：Unicodeコードポイント、XYWHなど

現代の平仮名のみ
変体仮名はなし

Unicode, x, y, w, h

U+3029, 512, 418, 56, 47

を



文字のバウンディングボックス（bounding box）の座標

ディープラーニング
によるくずし字認識
手法KuroNetを開発
1枚の画像の認識に約
1秒
条件が良ければ精度
95%



三てう殿に殿きたのかたならひておはし
ます御たいまいれりしううちよりまう
てたまへりくにくのしやうよりたうき
ぬぬのなともてまいれり御いそきのれう
にとてあやうす物かとりきぬなとお
ほく奉れたれはみくしけのする人御
まへまてはからひきたむそめくさ何くれの
としやうこのものともは一てう殿にもわかち
奉り給おはする事はなければ御かた
におほしなけきまくにきおとろかし

物体検出 (Object Detection)



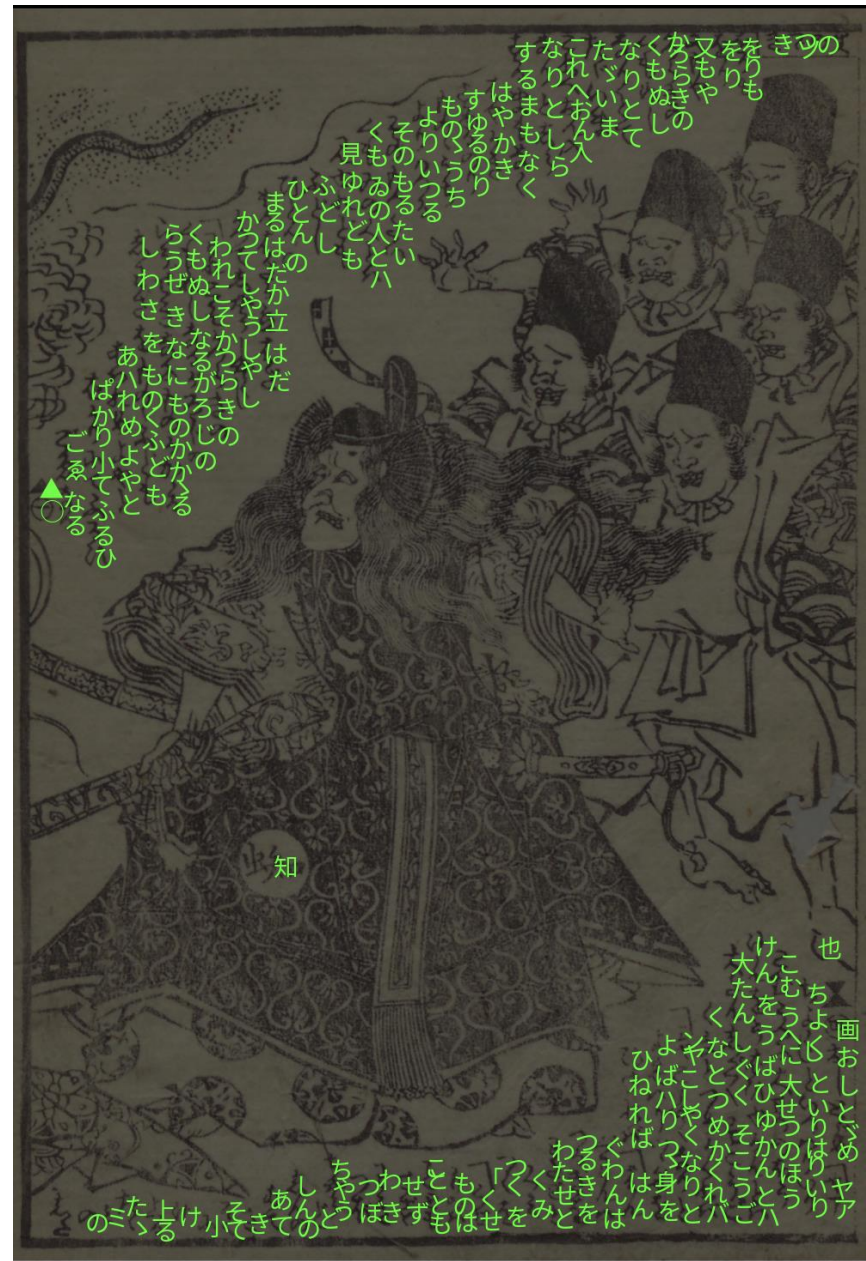
物体検出技術：顔検出や自動運転など幅広い応用がある。

くずし字 = オブジェクトとみなせば使える？

OCR処理順の逆転

- 通常のOCR：レイアウト解析→文字認識
- KuroNet：文字認識→レイアウト解析
 1. 処理順の逆転により、**レイアウト解析に依存しない文字認識**ができる
 2. 文字さえ認識できれば、**人間にとってレイアウト解析は難しくない**
 3. **木版印刷はレイアウトの自由度が高いため、機械学習により文字の並び順を完璧に推定することは難しい**

国立国会図書館蔵



kaggle くずし字認識コンペ

<http://codh.rois.ac.jp/competition/kaggle/>



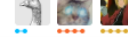
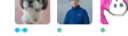
機械学習の研究者・技術者が
2000万人以上が登録する、**世界
最大のデータサイエンスプラッ
トフォームKaggle**にて、くずし
字認識コンペを開催

- **期間**：2019年7月19日～10月14日
- **参加チーム数**：293
- **参加者数**：338
- **結果提出回数**：2652

kaggle コンペの結果

上位の精度
は約95%

1. **さまざまな国の機械学習エンジニア、研究者が上位入賞**
2. **日本語やくずし字の知識がなくても開発可能**
3. **コンペの準備を滞りなく進めるには、**情報学者と人文学者の協働が不可欠****

#	Δpub	Team Name	Notbook	Team Members	Score	Entries
1	—	tascj			0.950	13
2	—	Konstantin Lopuhin			0.950	60
3	—	Kenji			0.944	161
4	▲1	YoudaoOCR			0.942	49
5	▼1	See--			0.940	42
6	—	abc			0.939	15
7	—	K_mat			0.934	20
8	—	t-hanya			0.920	21
9	—	Ollie, Nanashi, and Tom			0.910	35
10	—	Zenkei_R&D			0.903	144
11	—	masayai			0.903	12
12	▲5	Kirill Brodt (shad nsk)			0.901	4
13	▲1	James Day			0.901	33
14	▼1	NEU			0.900	54
15	▼3	s tatsuya			0.900	29

AIくずし字認識アプリ「みを」

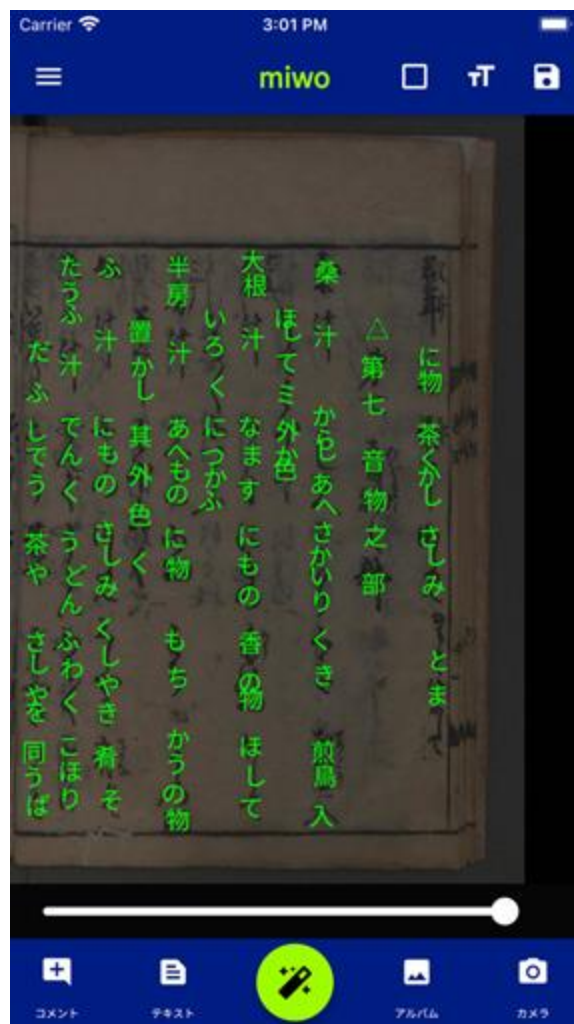
<http://codh.rois.ac.jp/miwo/>



- 『源氏物語』第14巻「みをつくし」（漣標）＝「みを（船の水路）を示すために立ててある杭」の意、「身を尽くし」の掛詞でもある
- 「みをつくし」が人々の水先案内となるように、「みを」アプリはくずし字資料を読むための道案内
- **2021年8月30日にiOS版とAndroid版を無料公開**
- アプリダウンロード数は約**20**万件、くずし字認識画像数は約**250**万件
- 開発：カラーヌワット・タリン氏（現在Sakana AI）

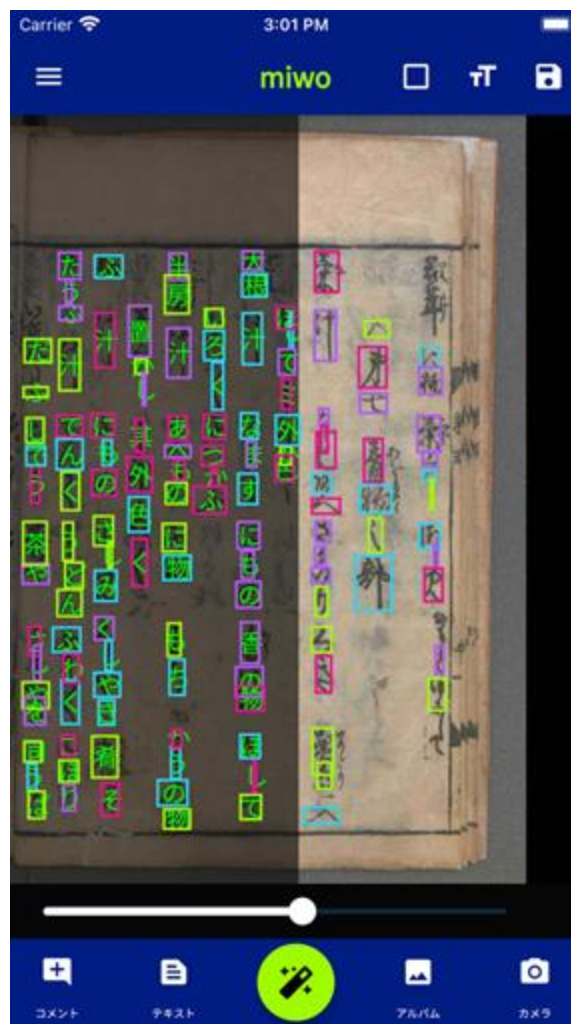


「みを」アプリのデザイン



くずし字認識

2024/10/26

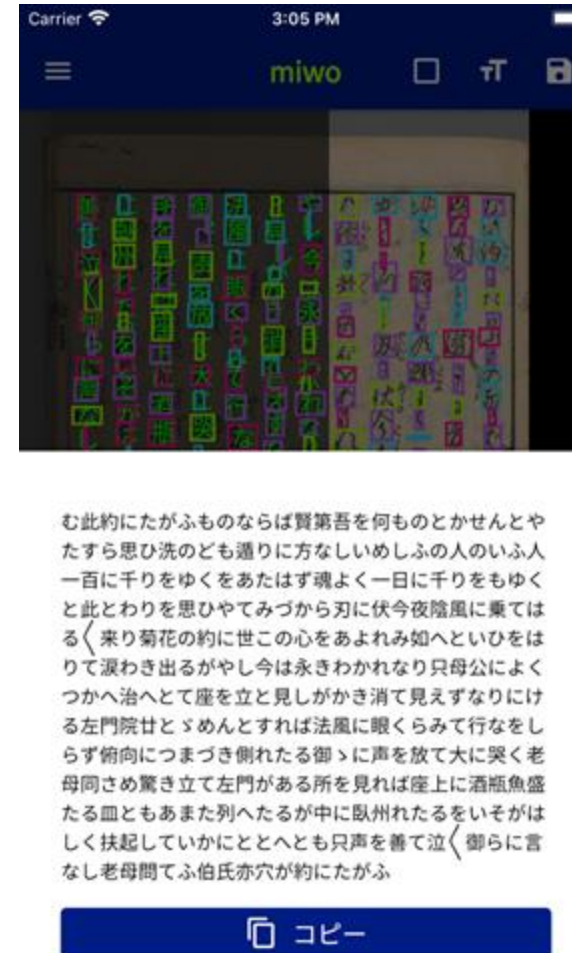


Bounding Box 表示

2024年度 軽井沢土曜懇話会 第3回



認識結果修正機能



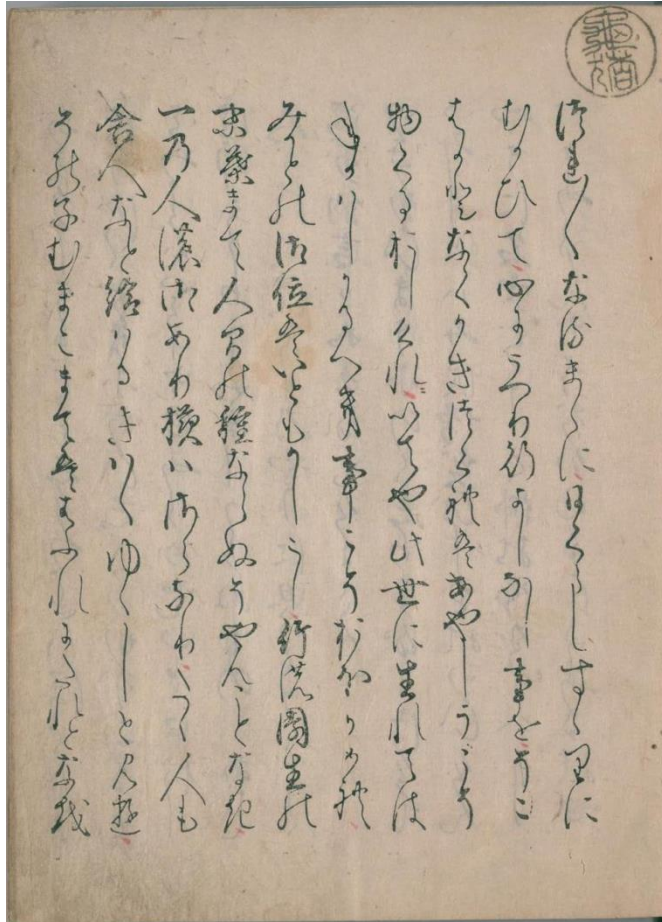
テキスト出力

くずし字認識とくずし字生成

1. **くずし字認識**：過去の文字→現代の文字に変換
2. **くずし字生成**：現代の文字→過去の文字に変換

1. **くずし字資料を読みたければ**、くずし字認識が役に立つ（例えば「**みを**」）
2. くずし字が読めない人でも、くずし字を生成できれば、**くずし字を使って学ぶ＋楽しむ**ことができる
3. **古活字データセット**を用いた**くずし字画像生成**

古活字版とは？



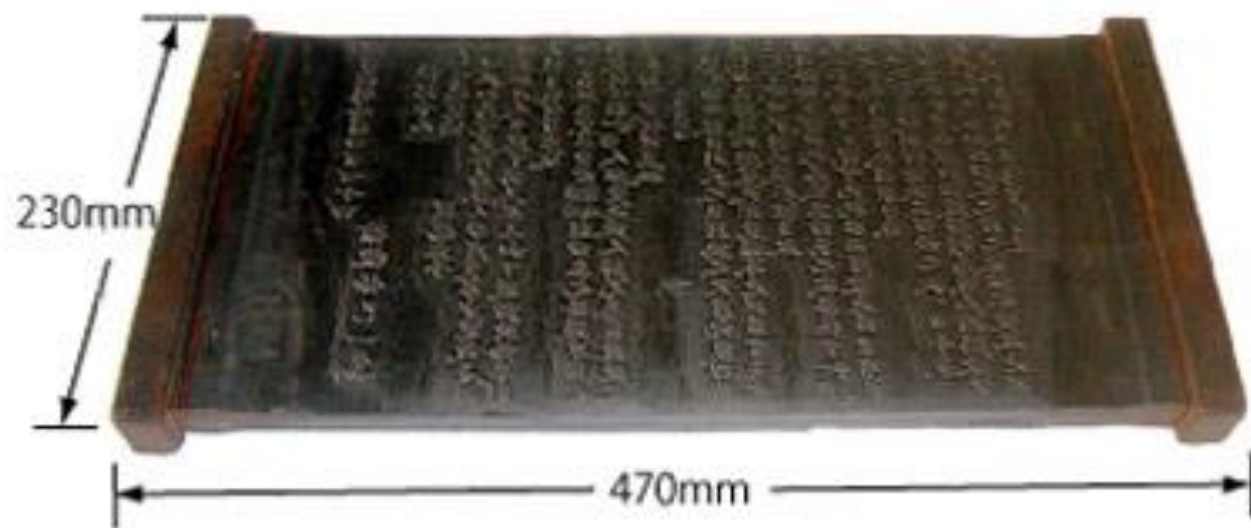
国立国会図書館デジタルコレクション

1. **古活字版**とは、16世紀末に西欧と朝鮮半島から伝えられた**活字印刷の技術に基づき**出版された本
2. 江戸時代初期の**角倉 素庵（すみのくら そあん）**は、京都嵯峨で出版業に関わった代表的存在
3. **古活字版「嵯峨本」**は、日本の出版史上もっとも美しい書物の一つであり、日本の書物文化の粹

日本の印刷技術

活版印刷

木版印刷



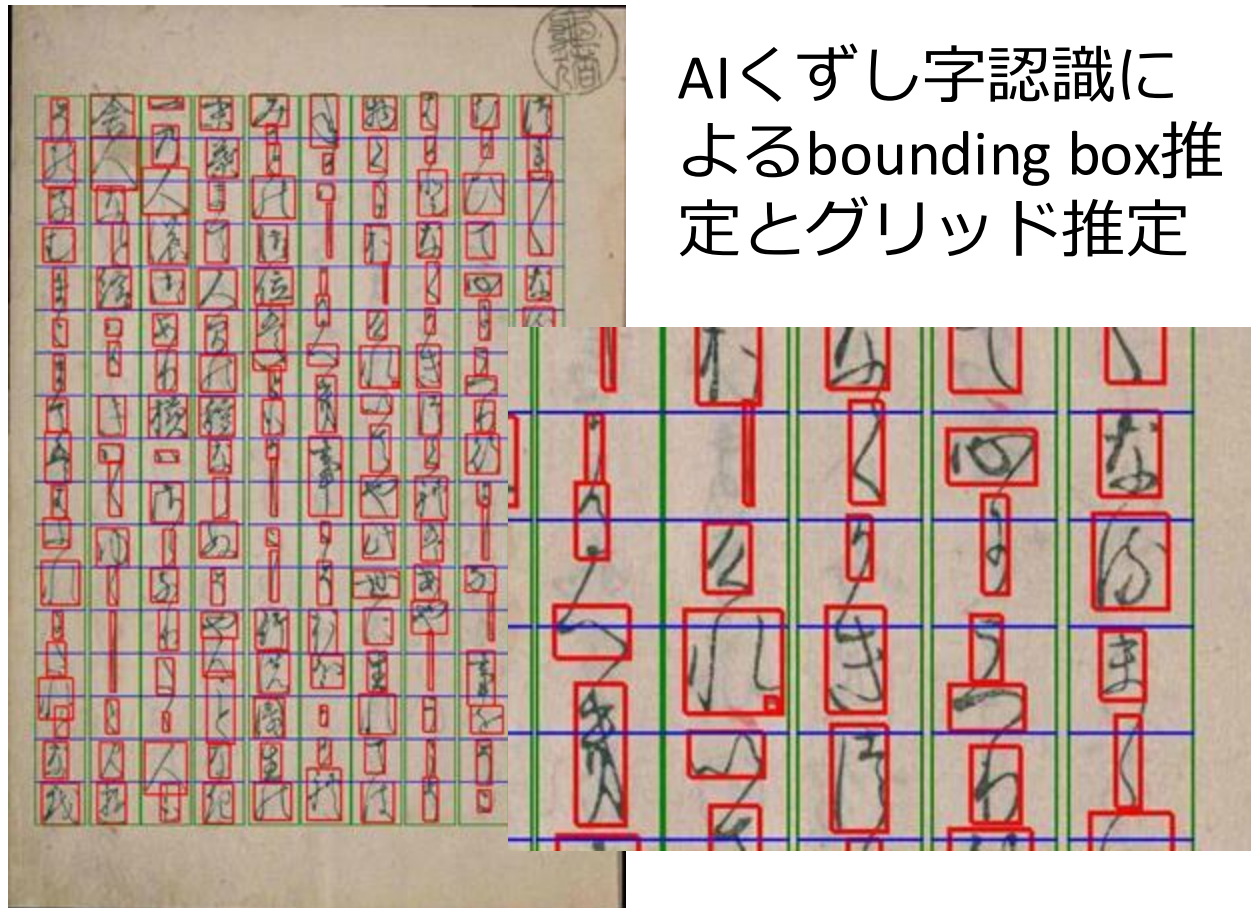
塙保己一史料館,
http://onkogakkai.com/aboutgunshoruijyu/gunshoruijyu_hangi/



印刷博物館, <https://www.printing-museum.org/collection/looking/15170.php>

古活字データセット

<http://codh.rois.ac.jp/omt/>



AIくずし字認識によるbounding box推定とグリッド推定

1. 古活字の実物がほとんど残存していないため、**印刷史として謎が多い**
2. 過去の活字を復元し、**デジタル組版**すれば、**任意の日本語をくずし字に変換できるのではないか？**

国立国会図書館デジタルコレクション

そあん (soan)

<http://codh.rois.ac.jp/soan/>

吾輩は猫である。名前はまだ無い。

どこで生れたかとうんと見当がつかぬ。何でも薄暗いじめじめした所でニャーニャー泣いていた事だけは記憶している。吾輩はここで始めて人間というものを見た。しかもあとで聞くとそれは書生という人間中で一番獰悪な種族であったそうだ。この書生というのは時々我々を捕えて煮て食うという話である。しかしその当時は何という考もなかったから別段恐しいとも思わなかった。

くずし字画像を生成！

サンプル:

吾輩は猫である

日本国憲法第九条

1. 任意の日本語テキストを、くずし字画像に変換可能
2. コラージュ技法：画像を断片化し、組み合わせ、合成する
3. 生成AI的要素はない

わはい、猫である。名前は、まふら
い。
とてまふらう電人ともあり、ほろ
ぬ。何てとうすゝゝい志めく
あふくやあふくやあふくや
事ふけりきむくゝゝゝ。わ
はい、あつて始りて人間という
ものを見た。あつてあつてあつと
うまひ書生という人間中て一冊
とうあつてあつてあつてあつて
だ。この書生という乃ハ町ノ我々
をとうえん煮て食うというあつ
てある。あつてその當時を何
んいうこうもあつてあつてあつて
しゝも思わあつた。

教育への利用

<http://codh.rois.ac.jp/edomi/learning/>

これまでのくずし字学習

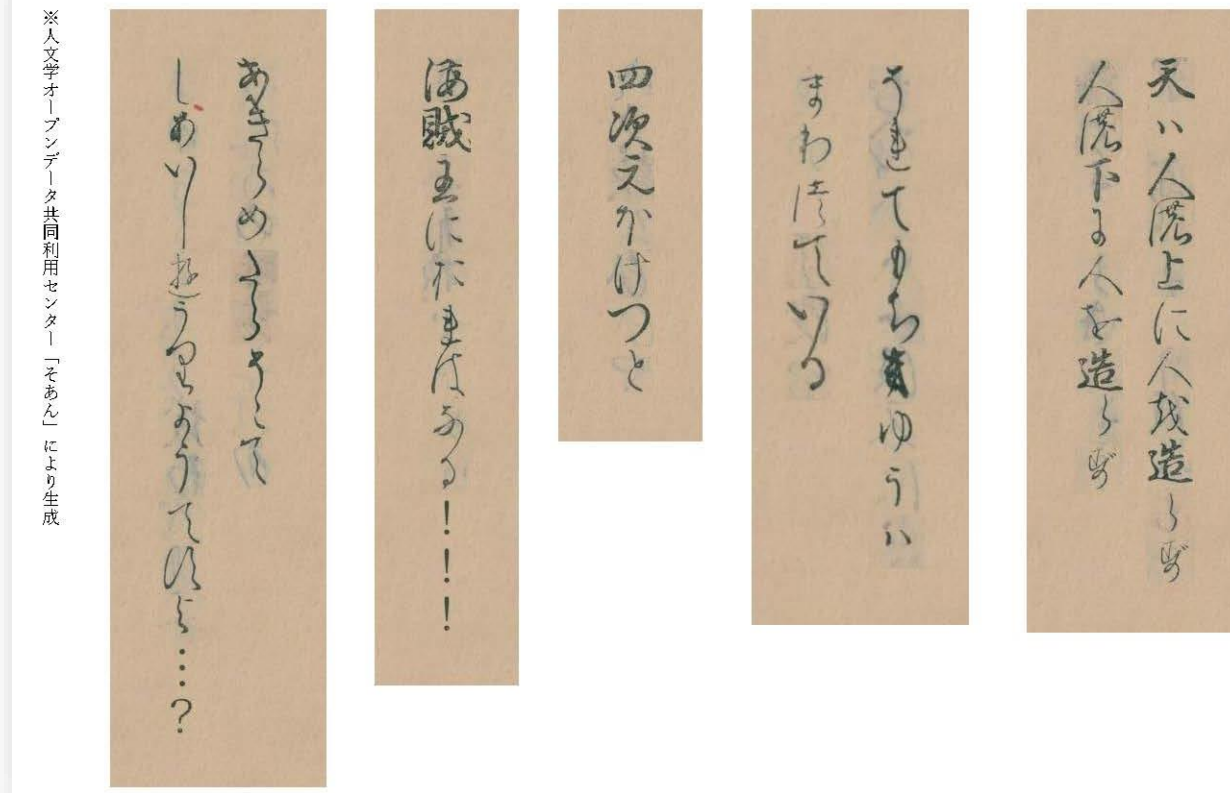
1. 文字と内容の両方を、同時に解読する必要がある
2. 内容の解読には、古文の文法や語彙も必要となる
3. 内容の理解には、背景知識が必要となる
4. 内容自体に興味を持てないと、すべての難関を乗り越える動機が高まらない

そあん（soan）を活用したくずし字学習

1. 文法や語彙、背景知識などを理解した、自分の好きなテキストを利用
2. 内容のギャップを縮め、文字の学習に集中できる

教室での利用

山形の古典籍をくずし字で読む

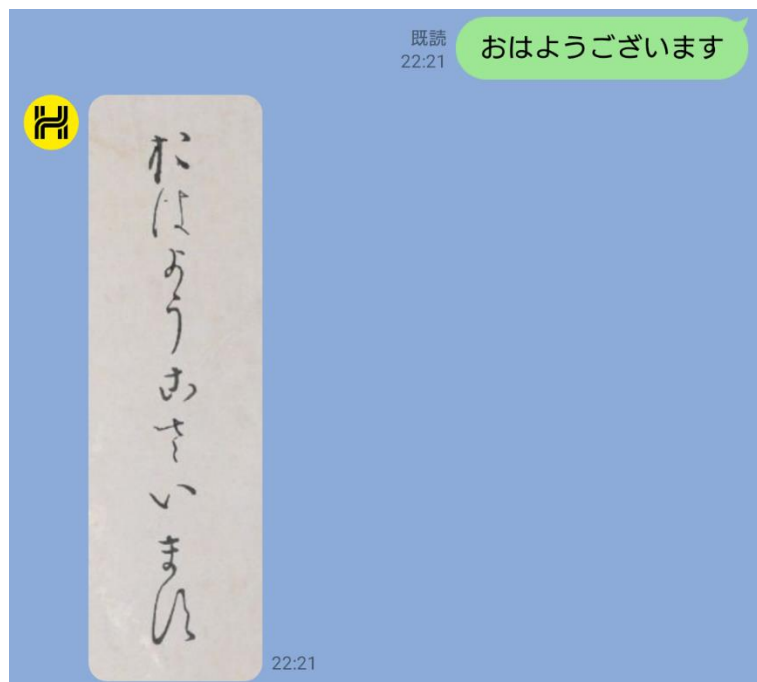


山形大学 生田慶穂氏 提供

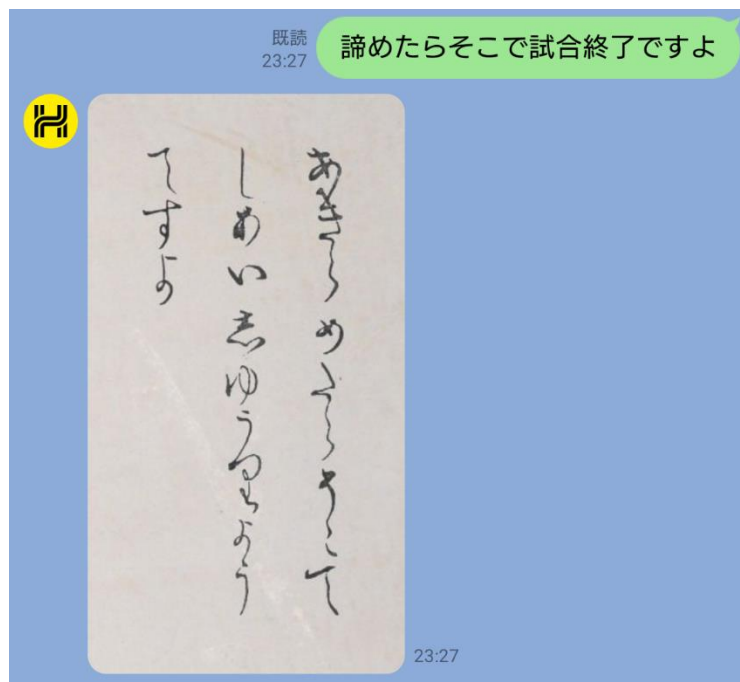
1. 大学1年生の授業で「そあん」生成教材を利用
2. 名言や漫画の有名なセリフを「そあん」で生成
3. 現代漫画のセリフによって、学生との距離を縮めることを期待
4. 感想：もっとくだらないものを読みたい

そあん (soan) ボット - LINEでくずし字 トーク

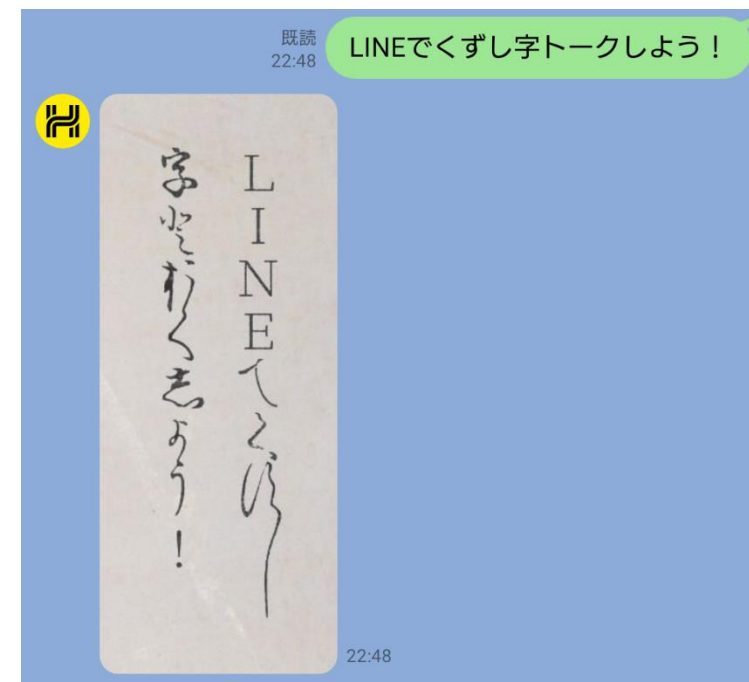
<http://codh.rois.ac.jp/soan/line/>



日常のあいさつや好きな言葉
など、何でもくずし字に変換
できます



漢字が表示できないときは、
読みを推定してひらがなを表
示します



漢字とひらがなだけでなく、
カタカナやアルファベットにも
対応します

再現と再生

- **再現**：過去の古活字版（例えば「嵯峨本」）を、できるだけ忠実に作り直す
- **再生**：過去の古活字版の遺産を、現代の人々が活用できる形式によみがえらせる
- **そあん（soan）の目標は再生**であり、そのために古活字版にない文字（アルファベット等）にも対応した
- **過去の遺産を再生させ現代で活用する方法論**：「江戸料理レシピデータセット」も同様の例

絵本江戸桜

«前 4 / 16 次»

☆ 目 下 上 右 左 ？



ICViewer | 日本古典籍データセット (国文研所蔵) CODH配信

AIチャット

チャット対象

本全体 ページリスト 画像領域

あなた

本全体の内容をまとめて

AI (gpt-4o)

「絵本江戸桜」は江戸の名所や風景を描いた絵本で、寛政7年正月に序文が書かれています。内容は、江戸桜の美しさや、浅草寺、寛永寺といった江戸の主要な寺社・名所が紹介されています。浅草寺の本尊である観音菩薩や、寛永寺の薬師如来など宗教的な要素も含まれています。また、新吉原の遊郭や両国橋の夏の納涼風景なども詳細に描写されています。全体として、江戸の風景や文化の豊かさが強調されています。



チャットを続ける



要約する

生成AI
チャット

IIIF画像
ビューア

「つくし」プロジェクト

<http://codh.rois.ac.jp/tsukushi/>

1. 日本古典籍（古文）に対して、現代日本語を用いた入力／出力を可能とする
2. 本の閲覧に使われてきたIIFの世界に大規模言語モデル（LLM）など生成AIの世界を接続
3. 本を閲覧しながら、LLMを基盤とする生成AIチャットと対話する仕組みを構築
4. 課金／非課金を含めたLLMの選択肢を用意し、日本古典籍に適したLLMの選択を支援

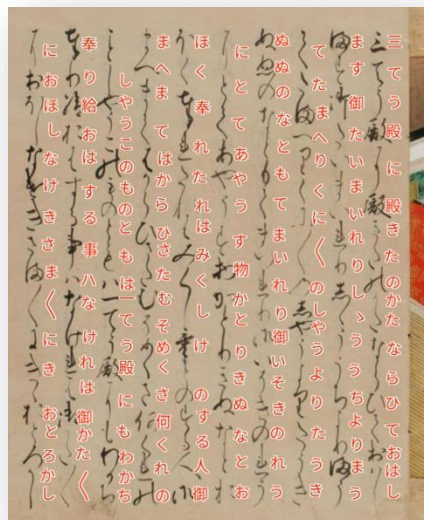
生成AIチャット

<http://codh.rois.ac.jp/tsukushi/ai-chat/>

プロンプト

この本に出てくるお寺の観光ガイドを作ってください。

IIIFの識別子のリスト



くずし字OCR

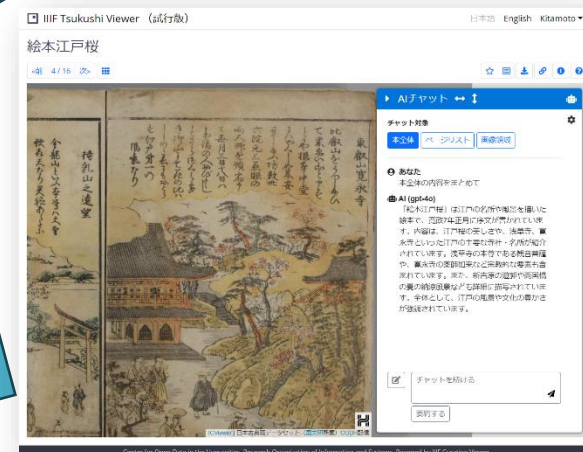
くずし字OCRしたテキスト (IIIFの単位でアクセス可能)

Content Service

大規模言語モデル
LLM

コンテキスト：古典籍の翻刻テキスト

プロンプト：ユーザが入力したテキスト





チャット対象



本全体

ページリスト

画像領域

● あなた

この本に出てくるお寺の観光ガイドを作ってください。

🎵 AI (gpt-4o)

東京のお寺観光ガイド

浄光寺（しょうこうじ）

日暮里に位置する浄光寺は、詠訪別当浄光寺の住庵として知られています。寺の隣には人丸大明神の社があり、境内からは筑波山や三河島村の風景が楽しめます。春には都人が花見を楽しむ絶景スポットです。

法明寺（ほうみょうじ）

雑司谷にある法明寺は、日蓮宗の寺院です。特に古松が茂る境内が魅力で、十月の会や紅葉見物に訪れる人々で賑わいます。名物の「やぶのそば」は風味豊かで評判です。

寛永寺（かんえいじ）

東叡山寛永寺は、比叡山を模して建立されました。根本中堂には薬師如来が安置され、毎月3日と18日には多くの参詣者が訪れます。桜の名所としても知られ、春には江戸第一の風景を楽しめます。



チャットを続ける



要約する

日本文化と生成AI

<https://sakana.ai/evo-ukiyoe/>

Sakana AI、立命館大学との共同研究により、「日本の美を学んだAI」を開発



Evo-Ukiyoe（画像生成）



Evo-Nishikie（画像カラー化）

テキストからの浮世絵生成

<https://sakana.ai/evo-ukiyoe/>



プロンプト：植物と花がある。
蝶が飛んでいる。

プロンプト：鶴が庭に立っている。
雪が降っている。



日本古典籍の挿絵カラー化

<http://codh.rois.ac.jp/tsukushi/evo-nishikie/>



浮世絵カラー化（**墨摺絵**⇒**錦絵**）をなぞった現代のカラー絵本

歴史ビッグデータの統合解析

<http://codh.rois.ac.jp/historical-big-data/>

過去のビッグデータを統合解析するための基盤技術进行研究



自然科学的データ

人文社会的データ

気候

地震

噴火

疫病

経済

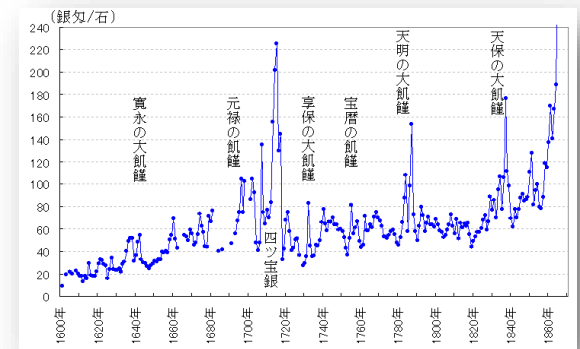
人口

政治

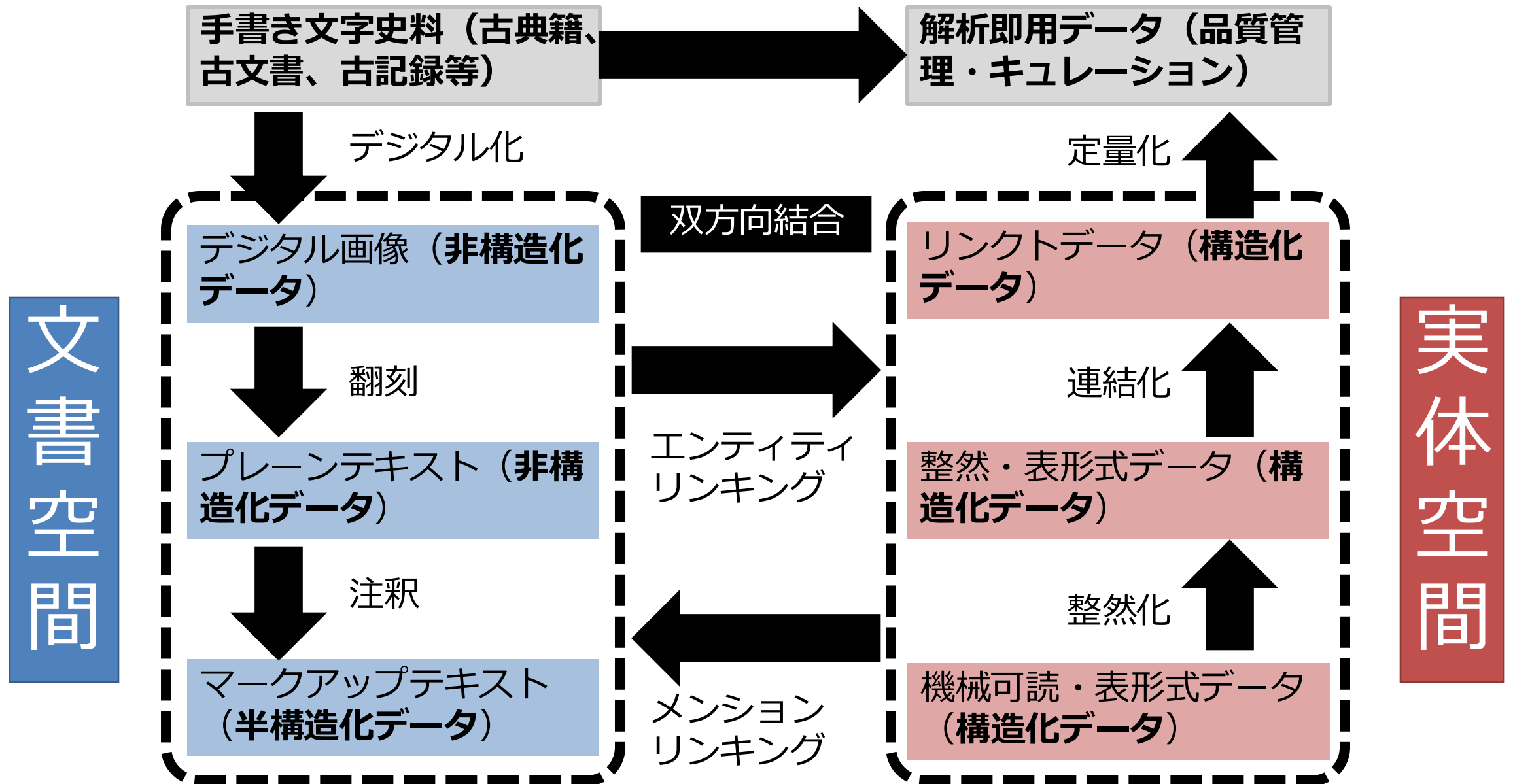
文化

データ構造化ワークフロー

歴史ビッグデータ研究基盤 (機械可読データ)



データ構造化ワークフロー



歴史地理データ

<https://geoshape.ex.nii.ac.jp/rekichime/>

1. 平凡社『日本歴史地名大系』
のデータを一部オープン化
2. 平凡社地図出版と協働し、位
置情報などの新たな価値を加
え、機械可読データ化
3. ジャパンナレッジIDを用いて、
オンラインDBとリンク
4. 出版社のオープン・クローズ
戦略の一つの事例



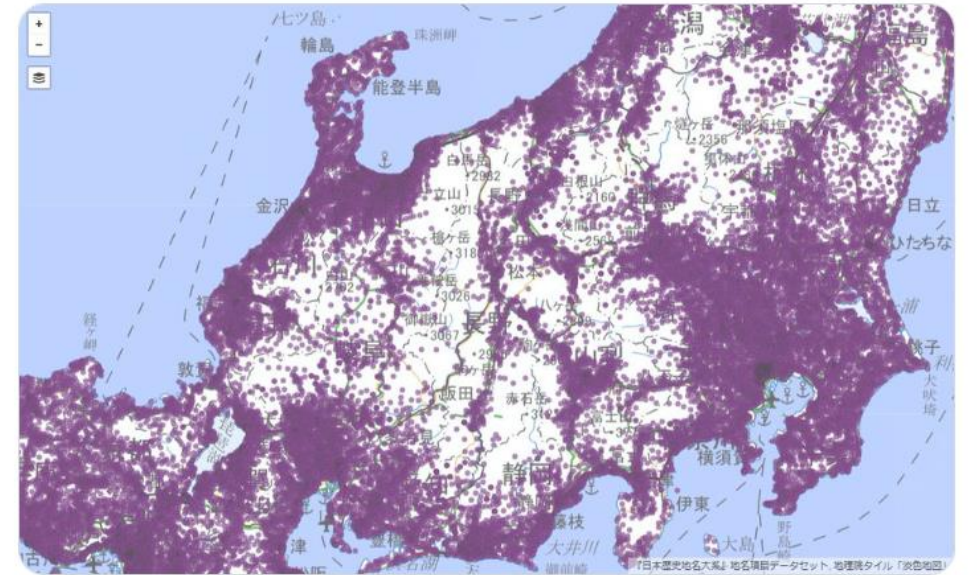
Center for Open Data in the Humanities (CODH) ✓

@rois_codh

歴史的地名の大規模オープンデータを公開しました。

codh.rois.ac.jp/news/#20231018

江戸時代の村までさかのぼる、79,502件の行政地名のデータです。平凡社地図出版と協働し、『日本歴史地名大系』の一部をオープン化しました。自分の先祖がどこに住んでいたかなど、地名から探っていくのも面白いですね。



午後5:21 · 2023年10月18日 · 79.3万 件の表示

📌 ポストのエンゲージメントを表示



5



3,422



7,491



2,691



江戸マップβ版

<http://codh.rois.ac.jp/edo-maps/>

29枚の古地図『江戸切絵図』から8722ヶ所の地名を抽出してデータベース化



番号	分類	現代語訳	翻刻	地図
2-001	 施設	幸橋御門	幸橋御門	拡大図
2-002	 施設	山下御門	山下御門	拡大図
2-003	 施設	数寄屋橋御門	数寄屋橋御門	拡大図
2-004	 施設	鍛冶橋御門	鍛冶橋御門	拡大図
2-005	 施設	呉服橋御門	呉服橋御門	拡大図
2-006	 地名	一石橋	一石橋	拡大図
2-007	 地名	出橋	出橋	拡大図
2-008	 町名	丸屋町	丸屋丁	拡大図

[2-296]
地名：磯辺大神宮（イソベ大神宮）
分類：寺社仏閣



千代田区

© 2020 ZENRIN

Google Earth

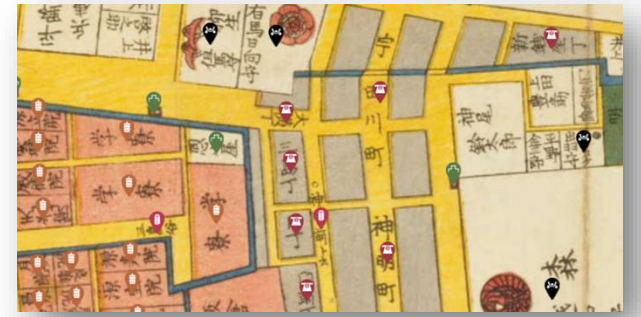
2024/10/26

2024年度 軽井沢土曜懇話会 第3回

1997

35° 40'52.36" N 139° 45'36.42" E 標高 2 m 高度 200 km

史料と地名のリンク



宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

固有表現認識

宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

曖昧性解消

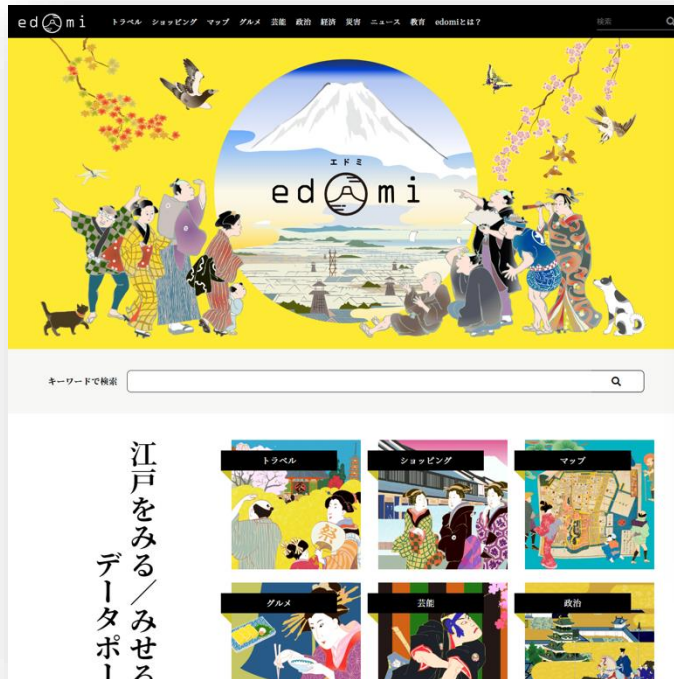
御江戸大地震大破并出火類焼場等書上之写（みんなで翻刻）

1. 江戸マップに出現する地名を、**地名リソース（エンティティのデータベース）**として整備
2. 史料の文字列から**固有表現（地名）**を抽出
3. **特定のエンティティとリンク**することで、実世界と紐づけ

原資料表記	江戸マップID	江戸マップ表記
宇田川丁	4-358	宇田川町
三島丁	4-290	三島丁
神明丁	4-294	神明町

edomi - 江戸をみる／みせるデータポータル

<http://codh.rois.ac.jp/edomi/>

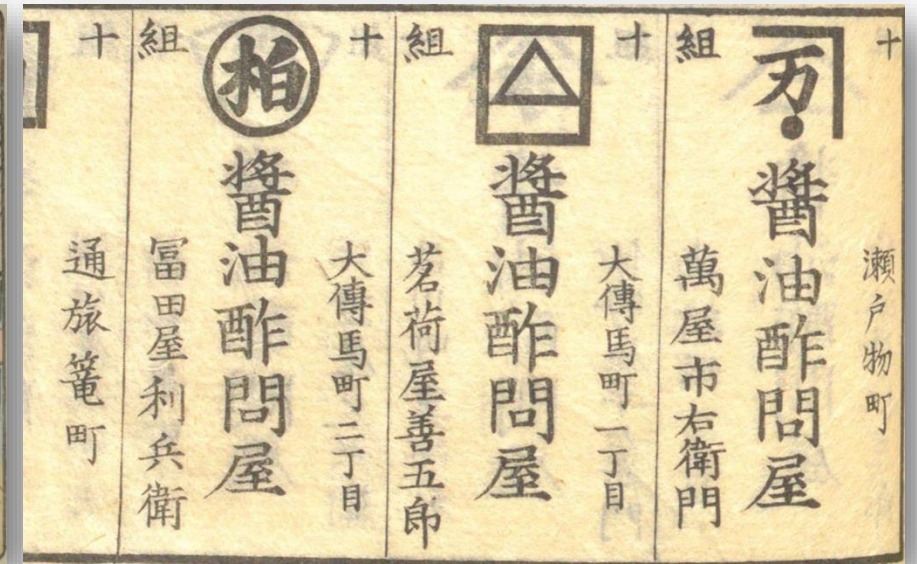


もし江戸時代に
ネットサービスが
あったら？



トラベル

江戸時代の旅行ガイドから作成した名所にIDと属性を付与

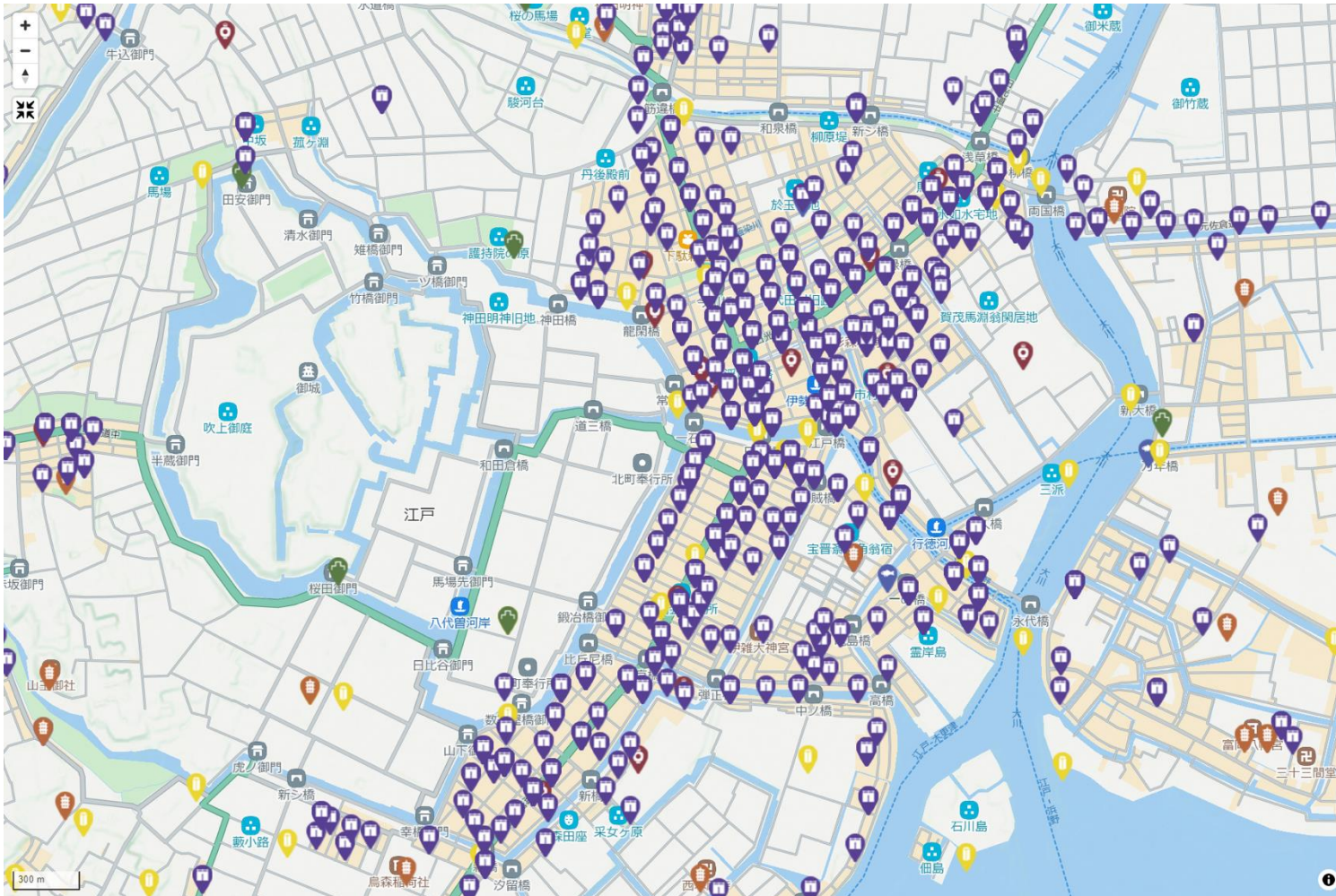


ショッピング

江戸時代の広告ガイドから作成した商店にIDと属性を付与

「れきちず」版江戸マップ

<http://codh.rois.ac.jp/edomi/map/>



1. れきちず：「現代風の地図デザイン」の歴史地図を閲覧できるサービス
2. 江戸切絵図の町家領域を彩色し、人口密集地を可視化

江戸の旅ルート

<http://codh.rois.ac.jp/edomi/route/>

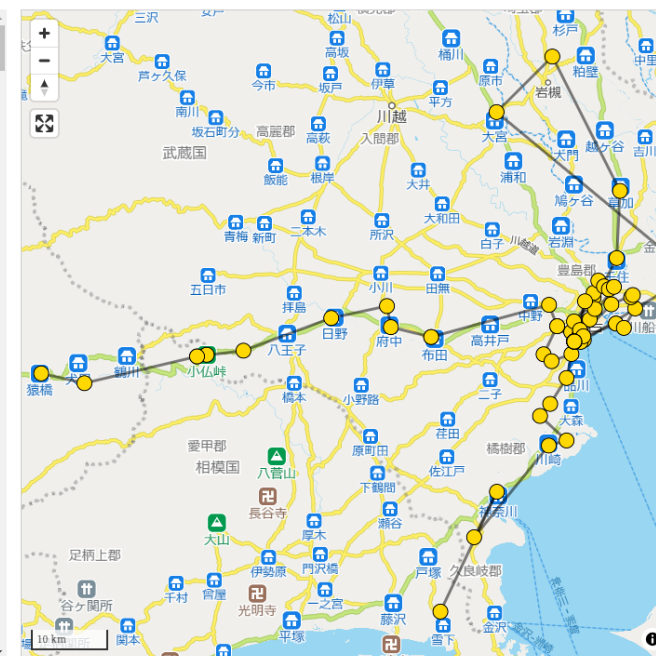
清河八郎『西遊草』

- 安政2年7月26日 [1] 堺の茶店 / 滞在
- 安政2年7月26日 [2] 保土ヶ谷 / 滞在
- 安政2年7月26日 [3] 神奈川宿 / 滞在
- 安政2年7月26日 [4] 台の茶屋 / 滞在
- 安政2年7月26日 [5] 立場の茶店 / 食事（午食）
- 安政2年7月26日 [6] 川崎宿 / 通過
- 安政2年7月26日 [7] 万年や / 通過
- 安政2年7月26日 [8] 大師堂 / 参拝
- 安政2年7月26日 [9] 羽根田の渡 / 通過
- 安政2年7月26日 [10] 羽根田村 / 通過
- 安政2年7月26日 [11] 羽根田弁天 / 食事（ハマグリ）
- 安政2年7月26日 [12] 大森 / 通過
- 安政2年7月26日 [13] 鈴ヶ森 / 通過
- 安政2年7月26日 [14] 品川 / 通過
- 安政2年7月26日 [15] 南品川 / 通過



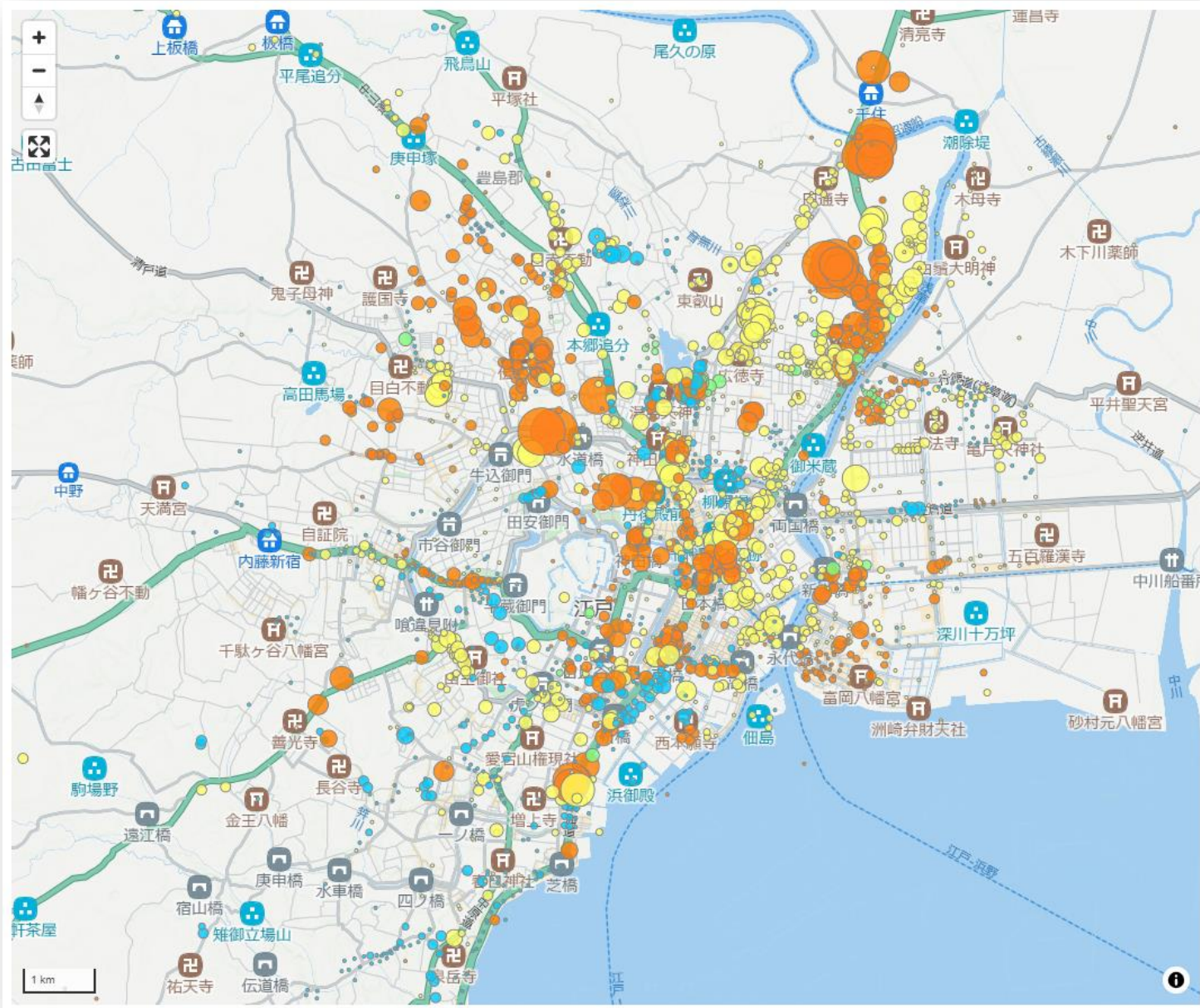
泉光院野田成亮『泉光院江戸旅日記』

- 文化13年2月7日 [1] 猿橋 / 渡る
- 文化13年2月7日 [2] 網土村 / 宿泊
- 文化13年2月8日 [3] 板橋（相模原市） / 宿泊
- 文化13年2月9日 [4] 小原、相模と武蔵の境の番所 / 通過
- 文化13年2月9日 [5] 小仏峠 / 引き返す
- 文化13年2月9日 [6] 板橋（相模原市） / 宿泊
- 文化13年2月10日 [7] 小仏峠 / 通過
- 文化13年2月10日 [8] 駒木（八王子市裏高尾駒木野）の幕府の番所 / 通過
- 文化13年2月11日 [9] 日野 / 宿泊
- 文化13年2月11日 [10] 武蔵国分寺 / 参拝
- 文化13年2月11日 [11] 六所明神 / 参拝
- 文化13年2月11日 [12] 石原（調布市上石原・下石原） / 宿泊



安政江戸地震 (1855年) 被害 記録分析

1. AIくずし字認識によるテキスト化
2. 地名認識・解決へのAI活用
3. テキストからの震度推定へのAI活用
4. ワークフローの各所でAIを活用



安政江戸地震被害記録マップ (試作)

日本文化と歴史ビッグデータ

1. 日本の過去の人々は、膨大な文書や書籍を大切に保存してきたが、現代の日本人はうまく活用できていない
2. AIを活用することで過去と現在の距離を縮め、現代の技術を過去に適用してアクセスを改善する
3. 歴史の解釈には複雑な文脈の理解が必要であり、AIに入れば答えが出るというものではない
4. 過去のデータを構造化し知識化することで、多くの人々が過去の教訓から学べるようにしたい

謝辞

1. AIくずし字認識やアプリ「みを」は、カラーヌワット・タリン氏（現在 Sakana AI）との共同研究です
2. くずし字生成は、国文学研究資料館「総合書物学」プロジェクトや本間 淳氏（フェリックス・スタイル）などとの共同研究です
3. Evo-Ukiyoe、Evo-NishikieはSakana AI、立命館大学との共同研究です
4. 「れきちず」は、株式会社MIERUNE（加藤創氏）が公開するオープンなサービスです
5. edomiは、鈴木親彦氏（現在群馬県立女子大学）との共同研究です
6. 「安政江戸地震被害記録」は、橋本雄太氏（国立歴史民俗博物館）、加納靖之氏、大邑潤三氏（東大地震研究所）との共同研究です