

背景

- 史的文字連携システムAPI
 - 合計6065字種、約32万件の古文書・古典籍の字形データを横断検索するシステムWEB API
 - 東京大学史料編纂所と奈良文化財研究所は2009年より構築・公開して、2020年より国文学研究資料館・国立国語研究所・京都大学人文科学研究所・台湾中央研究院歴史語言研究所が参加
 - JSON形式のデータを提供
- ポータルサイト「史的文字データベース連携検索システム」
 - 蓄積した字形画像を検索できる新たなとして公開している

問題点

ポータルサイトでは、検索手段として単文字検索のみのサポートであること、検索結果をカスタマイズできないことから、難読字形の確認といった目的での使用には課題がある。

解決方法

本発表では、上記のAPIを利用して、字種リストからの字形閲覧、異体字の変換、漢字部品での検索といった機能を供する検索ウェブアプリを開発する。

機能の実装

•異体字の変換

- APIでの検索はすでに編纂所異体字リストによって変換され
 - https://wwwap.hi.u-tokyo.ac.jp/ships/itaiji_list.jsp
- 簡体字と繁体字
 - OpenCC
- 漢字字体規範史データセット(HNG)

•字種リストからの字形閲覧

- 「かな」と「漢字」と分け
- 漢字を康熙字典部首

•漢字部品での検索

- 部品と画数で字種リストから検索

•検索結果のカスタマイズ

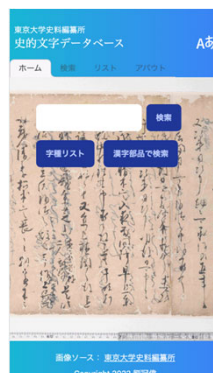
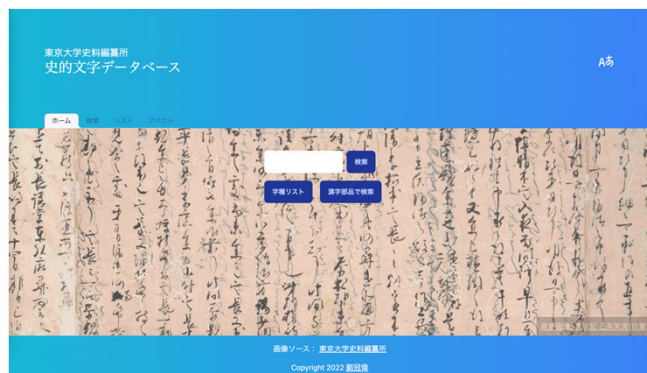
- 和暦西暦変換
 - HetimeのWEB APIを利用
- 年代でフィルタリング
 - 指定の年
 - 期間(実装予定)
- 編纂所の史料属性でフィルタリング
 - 職業(公家、武家など)
 - 史料区分(原本、影写本など)
- 結果の順序を時間順に表示可能
- IIIFでの字形画像表示
- 切り出し元画像へのリンク

• モバイル端末での利用

- 教育や調査の現場に利用しやすい
- 多言語対応
- 日中英

結果と考察

DEMO URL: <https://hi-kanji.lab.hi.u-tokyo.ac.jp/>



•考察

- 検索:
 - 異体字変換
 - 部品画数での漢字検索
 - 表示:
 - 西暦変換
 - 属性フィルタリング
 - 表示順序
 - 携帯対応
- WEB APIとして言語資源を利用例

•課題

- 編纂所以外の史的文字API
- 史的文字API以外の字形データ
- 複数APIの横断検索