Vol.2019-CH-120 No.2
2019/5/11

漢字字体規範史データセット用従来型 UI 再生の試み

守岡 知彦[1]　劉 冠偉[2]　高田 智和[3]

概要：2015 年 4 月頃に停止した漢字字体規範史データベース (HNG) の代替サービスとしては 2015 年 11 月に公開された「CHISE-IDS HNG 漢字検索」が存在するが、元々の HNG とユーザーインターフェース (UI) が大きく異なるため旧来の UI が使いたいというユーザーの声が寄せられていた。このため、我々は「HNG 単字検索」と呼ぶ従来型 UI を提供する検索サービスを開発した。この検索サービスは従来風の操作性や表示を再現するだけでなく、検索結果に対するパーマリンクやレスポンシブ Web デザインを実現することでスマートフォンやタブレットを含んだ現代的な Web 環境における快適な操作性の実現を目指した。また、EgT/ CHISE-wiki 上の紙カード画像及び文字情報へのリンクを通じて CHISE ベースの HNG の機能や Unihan / GlyphWiki 等の国際的に利用されている文字データベースとの連携も実現している。本発表ではこの HNG 単字検索の概要について概説する。

1. はじめに

漢字字体規範史データベース (Hanzi Normative Glyphs; HNG) [1] は時代や地域毎の漢字字体の標準の存在とその変遷を明らかにすることを目的に構築された漢字のグリフデータベースである。その前身は石塚晴通氏が 30 年程前から作成を続けてきた字体資料（「石塚漢字字体資料」と呼ぶ）である。「石塚漢字字体資料」は紙カードで整理されていたが、15 年程前から電子化が開始され、2005 年から豊島正之氏の管理のもとで Web 上での検索サービスの公開が始まった。

HNG は収録する資料に含まれる字形用例を字体に分類し、その代表字形を字種（抽象文字）によって管理し、字種粒度によるソースをまたいだ串刺し検索・一覧表示を実現することにより、文字（字種単位）での字体の変遷を把握可能にしている。また、標準・規範的な度合が強い公的な写本・版本（および、石刻文字）を中心に私的な写本・版本等も比較のために収録することで、楷書の字体規範の時代的・地域的変遷と異化の全体像を把握することを可能にしている。特に、敦煌本を始めとする唐代以前の中国古写本と、奈良・平安時代の日本古写本を通して、初唐の標準字体が日本の標準字体として移入・定着する様相を精緻に記述する基盤を提供している。また、開成石經の字体が宋版を通じて受容されることによって初唐の標準字体とは異なる字体が新たな規範字体として定着する様子を描写している。

HNG が公開から 10 周年を迎えた 2015 年の 4 月頃[*1]、HNG はサービスを停止した。2015 年 11 月には HNG 公開 10 周年を記念するイベントが企画されており、その発表者は HNG が利用できない状況で HNG について語らざるを得ないという状況に陥った。

そこで、著者らは HNG の古いバックアップデータを元に CHISE [2] を利用した HNG の代替サービスを開発し、2015 年秋頃から CHISE-wiki [3] 上での HNG 字体資料の公開を始め、11 月には「CHISE-IDS HNG 漢字検索」を公開した。[4] また、HNG 公開 10 周年イベントの場で、ユーザーから「石塚漢字字体資料」の紙カード画像を公開して欲しいという要望があり、石塚晴通氏他関係者が一般公開することで合意したため、2016 年 4 月に CHISE-wiki に石塚漢字字体資料の紙カード画像の表示機能を追加し、2016 年 6 月には IIIF Image API を利用した紙カード画像と開成石經拓本画像の配信、及び、これを利用した京大人文研所蔵の開成石經画像と HNG の開成石經データの比較表示機能も追加した。[5]

こうした取組みにより、48 資料に限られるものの HNG の情報が再びインターネット上から自由に利用できるようになり、また、「CHISE-IDS HNG 漢字検索」を用いた部品単位での検索や比較、紙カード画像に戻っての再検討、

[1] 京都大学人文科学研究所
Institute for Research in Humanities, Kyoto University
[2] 北海道大学文学院／学術振興会特別研究員-DC2
Graduate School of Humanities and Human Sciences, Hokkaido University / JSPS Research Fellowship for Young Scientists
[3] 国立国語研究所
National Institute for Japanese Language and Linguistics

[*1] 正確な日時は良く判らない。

© 2019 Information Processing Society of Japan

HNG の開成石經の拓本よりも良い拓とされる京大人文研所蔵の開成石經画像との連携機能といった従来にはなかった高度な機能が追加され、CHISE やそれと相互リンクしている GlyphWiki や UniHan データベースといった新たな導線ももたらされた。その一方で、旧来のユーザーインターフェース (UI) が使いたいユーザーの要望を十分にかなえたものとはなっていなかった。[6]

そこで、我々は停止前の「漢字字体規範史データベース」(旧 HNG 検索) に近い操作性や画面表示を持ちつつ、現在の Web 環境に適合し、また、CHISE との統合によってもたらされた新たな機能も利用可能な漢字字体規範史データセット用の検索サービスとして「漢字字体規範史データセット単字検索」(略称「HNG 単字検索」)

`https://search.hng-data.org/`

を開発した。本稿ではその概要について概説する。

2. 漢字字体規範史データセット

現在、我々は石塚漢字字体資料および漢字字体規範史データベースの内容を長期にわたり安定的に利用可能にするために、これらの情報や関連する情報を調査・整理し、Git によって版管理されたデータセットにまとめる取組みを行っている。これを「漢字字体規範史データセット」[7] (「HNG データセット」と略す) と呼ぶ。

このデータセットの Git リポジトリ [8] は前述の CHISE を利用した HNG 関連サービスの開発時に作成したものをベースにしている。これは 2006 年度時点でのバックアップデータをベースに開発が始まったため、48 資料しか収録できていなかった [9] が、 最終的に公開上の問題がないと考えられる 63 資料を収録する予定である。

2019 年 4 月現在、60 資料を収録しており、最新版だけでなく、過去の版の情報もブランチやタグ等を利用して参照可能にすることを計画している。

3. 要求仕様

3.1 システムに関するもの

旧来風の UI を持った検索サービスを開発するにあたり、単に従来風の操作性や表示を再現するだけではなく、現代的な Web 技術を用い現在の Web 環境と調和した実装を実現することを目指した。また、CHISE と連係することで、紙カード画像や拓本画像の利用や CHISE の文字情報、Unihan / GlyphWiki 等との連携といった CHISE-wiki (EgT [10]) 上で実現された新たな機能を利用することも目指した。

具体的には、

(1) 検索結果に対するパーマリンクの実現
(2) レスポンシブ Web デザインの実現
(3) 石塚漢字字体資料の紙カード画像 (EgT) へのリンク
(4) CHISE-wiki 上の HNG 例示字形オブジェクトの頁へのリンク
(5) 見出し字 (UCS 抽象文字)、大字典番号 (字種)、大漢和番号 (例示字体) に対応する CHISE-wiki 上の文字オブジェクトの頁へのリンク

を要求仕様とした。

3.2 データに関するもの

従来の CHISE ベースの HNG 関連サービスは 48 資料のみをサポートしていたが、HNG 単字検索では HNG データセットが収録対象としている 63 資料全てを対象とすることにした。

また、旧 HNG 検索では見出し字および異体字の符号化において BMP の範囲内でしか UCS 符号化を行っていなかったが、漢字字体規範史データセットでは統合漢字拡張 F までの現時点で利用可能な UCS に収録された統合漢字の全てをサポートすることを目指し、統合漢字拡張 B 以降へのマッピング情報の追加を目指すことにした。これに伴い、HNG 単字検索も旧 HNG 検索では利用できなかった拡張 B 以降の統合漢字をサポートする。

4. 実装

4.1 システム

HNG 単字検索のサーバー側の実装には PHP で書かれたオープンソースの Web アプリケーションフレームワークである Laravel を用いた。また、フロントエンドにはオープンソースの JavaScript フレームワークである Vue.js 及び Vue.js 用のマテリアルデザインフレームワークである Vuetify を用いた。図 1 にその構成図を示す。

4.2 API

HNG 単字検索サーバーは次のような API を提供している。

4.2.1 全用例検索

`/api/v1/search/character/`c`/all`

漢字 c を検索し、HNG データセットにおける c の全用例の情報を JSON 形式で返す。

例： `https://search.hng-data.org/api/v1/search/character/学/all`

→


```
{
    "chars": {
        "id": 2293,
        "entry": "學",
        "variants": "学孝",
        "djt": "02293",
        "dkw": "M07033",
        "ucs": "5B78",
        "created_at": "2019-04-08 14:33:44",
        "updated_at": "2019-04-08 14:33:44"
```


© 2019 Information Processing Society of Japan

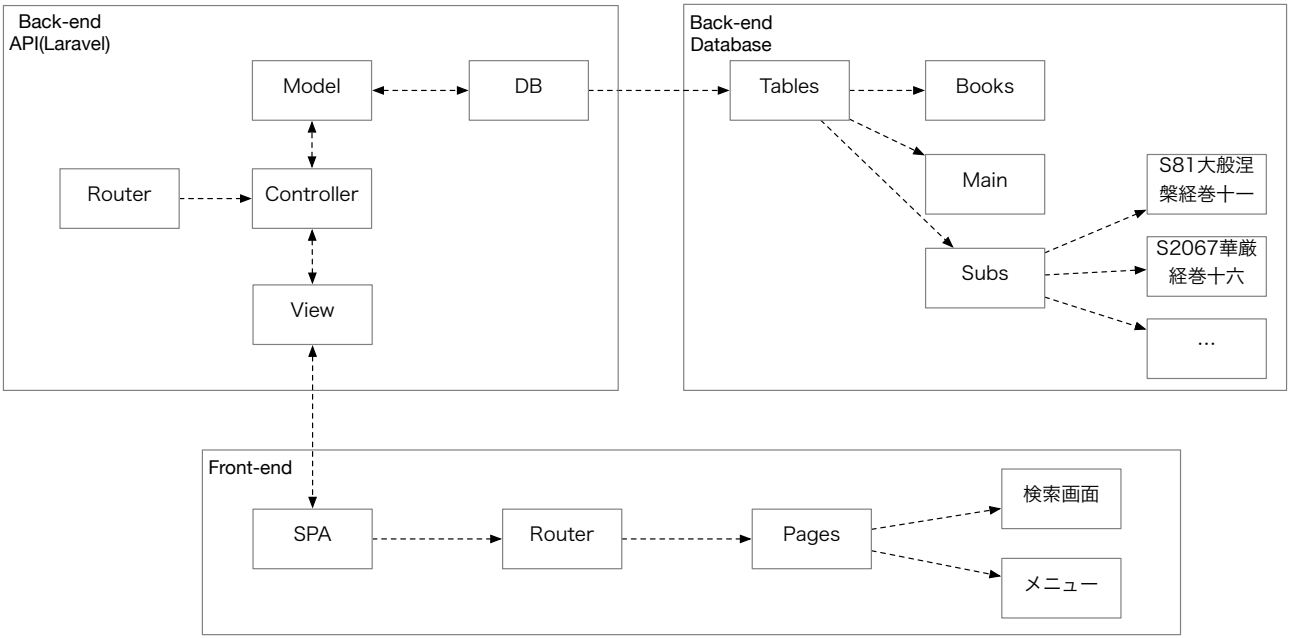


図 1 HNG 単字検索の構成


```
},
"CM": {
    "dng": [
        {
            "id": "02293",
            "entry": "學",
            "variants": "学斈",
            "djt": "02293",
            "dkw": "M07033",
            "ucs": "5B78",
            "created_at": "2019-04-08 14:33:45",
            "updated_at": "2019-04-08 14:33:45",
            "glyph_num": "0472",
            "glyphs": "1",
            "occerrences": "2",
            "chisechar": "dng=4720",
            "hngcard": "3-472"
        }
    ],
    "khm": [
        {
            "id": "02293",
            "entry": "學",
            "variants": "学斈",
            "djt": "02293",
            "dkw": "M07033",
            "ucs": "5B78",
            "created_at": "2019-04-08 14:33:47",
            "updated_at": "2019-04-08 14:33:47",
            "glyph_num": "0027",
            "glyphs": "0",
            "occerrences": "1",
            "chisechar": "khm=270",
            "hngcard": "11-27"
        }
    ],
    "hok": [
        {
            "id": "02293",
            "entry": "學",
            "variants": "学斈",
```


（以下略）

4.2.2 指定した資料内における用例の検索

/api/v1/search/character/*c*/*s*

資料 ID *s* で指定する文献において漢字 *c* を検索し、その結果を JSON 形式で返す。

s で指定する文献の ID は次の booklist API で参照可能。

例： https://search.hng-data.org/api/v1/search/character/学/kar

→


```
[
    {
        "id": "02293",
        "entry": "學",
        "variants": "学斈",
        "djt": "02293",
        "dkw": "M07033",
        "ucs": "5B78",
        "created_at": "2019-04-08 14:33:47",
        "updated_at": "2019-04-08 14:33:47",
        "glyph_num": "1429a",
        "glyphs": "2",
        "occerrences": "50"
    },
    {
        "id": "02293b",
```


© 2019 Information Processing Society of Japan

```
        "entry": "學",
        "variants": "学孛",
        "djt": "02293",
        "dkw": "M07033",
        "ucs": "5B78",
        "created_at": "2019-04-08 14:33:47",
        "updated_at": "2019-04-08 14:33:47",
        "glyph_num": "1429b",
        "glyphs": "2",
        "occerrences": "1"
    }

]
```


4.2.3 資料一覧の取得

`/api/v1/booklist`

HNG データセットの資料一覧を JSON 形式で返す。

例： `https://search.hng-data.org/api/v1/booklist`

→


```
[
    {
        "no": "No. (code)",
        "id": "ID",
        "kubun": "区分",
        "name": "文献名",
        "type": "地域種別",
        "short_name": "略称",
        "date_order": "作成順",
        "date": "作成年代",
        "normal_type": "標準",
        "koushi": "公／私",
        "shozo": "所蔵",
        "created_at": "2019-04-08 14:33:45",
        "updated_at": "2019-04-08 14:33:45"
    },
    {
        "no": "1 (H03)",
        "id": "dng",
        "kubun": "南北朝写本",
        "name": "S81 大般涅槃経巻十一",
        "type": "CM",
        "short_name": "S81",
        "date_order": "1",
        "date": "506",
        "normal_type": "標準",
        "koushi": "",
        "shozo": "BL",
        "created_at": "2019-04-08 14:33:45",
        "updated_at": "2019-04-08 14:33:45"
    },
    {
        "no": "2 (H02)",
        "id": "keg",
        "kubun": "南北朝写本",
        "name": "S2067 華厳経巻十六",
        "type": "CM",
        "short_name": "S2067",
        "date_order": "2",
        "date": "513",
        "normal_type": "標準",
        "koushi": "公的",
        "shozo": "BL",
        "created_at": "2019-04-08 14:33:45",
        "updated_at": "2019-04-08 14:33:45"
    },
    {
        "no": "3 (H01)",
        "id": "jou",
        "kubun": "南北朝写本",
        "name": "P2179 誠実論巻八",
        "type": "CM",
        "short_name": "P2179",
（以下略）
```


4.2.4 Web 検索の実行

`/search/`c

漢字 c を検索し、結果の Web ページ (HTML) を返す。

例： `https://search.hng-data.org/search/学`

→ 図 3 (cf. 5.1 節)

4.3 データセット

2019 年 4 月現在ではまだ HNG データセットに収録されていないものも含め 63 資料を先行して収録した。

また、CHISE ベースの HNG 関連サービスにおいても資料の追加作業を行っており、2019 年 4 月現在、60 資料を収録している。この 60 資料に関しては紙カードの画像も表示可能である。

5. Web UI

「漢字字体規範史データセット単字検索」（略称「HNG 単字検索」）のトップページである

`https://search.hng-data.org/`

にアクセスすると図 2 のような検索窓が一つだけあるシンプルな画面が現れる。HNG 単字検索の使い方は基本的にはこの検索窓に探したい 1 文字を入力するだけという極めて単純なものである。

5.1 検索結果の URL

例えば、検索窓に「学」を入れて検索ボタンを押すと図 3 のような検索結果を表示する画面が現れ HNG の各用例が表示される。

この画面は指定した検索対象である文字を c とする時、これに対応した固有の URL

© 2019 Information Processing Society of Japan



図 2 HNG 単字検索のトップ画面

`https://search.hng-data.org/search/`*c*
を持つ。

例えば、「学」を指定した場合、その検索結果のパーマリンクは

`https://search.hng-data.org/search/学`
となる。

5.2 検索結果画面の構成

検索結果画面は大きく分けてヘッダー、見出し字情報、用例情報、フッターからなる。

5.2.1 ヘッダー部

ヘッダーはトップページと同様なもので、左右のアイコンからメニューを出すことができる。

5.2.2 見出し字情報表示部

見出し字情報は検索した文字に対応する見出し字の情報を簡潔に表示する部分で、「見出し字」、「大字典番号」、「大漢和番号」の値部分はそれぞれ CHISE-wiki における見出し字に対応する UCS 抽象文字オブジェクト、大字典字種（超抽象文字）オブジェクト、大漢和の例示字体オブジェクトのページに対するリンクとなっており、これらのリンクをたどることによってより詳細な情報を見ることもできる。この内、大字典字種オブジェクトは HNG 単字検索に対応するために大字典データベース [11] [12] をもとに今回新設したものである。

5.2.3 用例情報表示部

用例情報は HNG における用例を表示する部分で、「中国写本」、「中国版本」、「日本写本」、「日本版本」、「韓国資料」、「その他資料」の 6 段に分けて HNG データセットに収録された各資料での用例が表示される。各資料の項目は上段に略称、中央に代表字形、下段に用例数が表示される。また、代表字形の画像をクリックすると CHISE-wiki における代表字形オブジェクトのページに飛ぶことができる。同様に、用例数をクリックすると EgT における石塚漢字自体資料の紙カードのページに飛ぶことができる。これらのページを見ることにより例示字形の情報や紙カードの情報を確認することができる。

5.2.4 フッター部

フッターにはトップページのと同様な検索窓と漢字字体規範史データセットの Web サイトへのリンクがある。

6. おわりに

「漢字字体規範史データセット単字検索」（略称「HNG 単字検索」）について概説した。

漢字字体規範史データベース（旧 HNG 検索） の代替サービスとしては CHISE-wiki (EgT) を利用した HNG 関連サービスが存在するが、ユーザーインターフェース (UI) が大きく異なるため従来風の UI が使いたいという声があった。「HNG 単字検索」はこうした声に答えるべく開発した Web サービスであるが、単に旧 HNG 検索に似た UI を提供するだけでなく、利用頻度の少ない機能を削ることでよりシンプルにするとともに、検索結果に対するパーマリンクやレスポンシブ Web デザインを実現することでスマートフォンやタブレットを含んだ現代的な Web 環境に適合した Web サービスとなっている。また、EgT/ CHISE-wiki 上の紙カード画像及び文字情報へのリンクを通じて CHISE ベースの HNG の機能や Unihan / GlyphWiki 等の国際的に利用されている文字データベースとの連携も実現している。

「HNG 単字検索」の実現はある程度のユーザーを抱えながらもサービスを停止してしまった人文系データベースの再生に関するケーススタディーとしての意義もあると思われる。データベースの長期保存にはデータの保存も重要であるが、『データの操作』や UI のようなデータに関する計算やプログラムといったものの保存も重要であると考えられる。こうしたものの保存としては仮想計算機などを用いてプログラムやシステムそのものを保存するというアプローチも考えられるが、ネットワーク上のサービスの場合、セキュリティーの問題やユーザー側の環境の問題などからこうしたアプローチに基づいて実用的なサービスを保存することは難しいと思われる。特に、UI がからむものの場合、この問題は顕著であろう。こうしたことを考えれば、各時代の計算機環境に適合するようにシステムを作り変えていくことが重要であると考えられる。そのためには、データと処理の関係やスクリーンショット、あるいは、それらの背景となる考え方を記録し保存することが重要であるといえる。本稿は「HNG 単字検索」を紹介するとともに、今後の作り直しの際の一助となることを期待して記述したものである。

© 2019 Information Processing Society of Japan



図 3 HNG 単字検索の検索結果の例（学）

謝辞

本研究の遂行にあたり科研費基盤研究 (C)「字体記述の精密化手法の確立による歴史的漢字字体情報アーカイブズ構築」（課題番号：18K00611）の助成を得た。また、石塚晴通氏、豊島正之氏、池田証壽氏、斎木正直氏、及び、北海道大学文学研究科池田証壽研究室の諸氏に感謝する。

参考文献

[1] 石塚晴通，池田証寿，岡墻裕剛：漢字字体規範データベースとその応用，東洋学へのコンピューター利用第 17 回研究セミナー，pp. 53–63 (2006).
[2] Morioka, T.: Multiple-policy Character Annotation based on CHISE, *Journal of the Japanese Association for Digital Humanities*, Vol. 1, No. 1, pp. 86–106 (2015).
[3] 守岡知彦：CHISE のセマンティック Wiki 化の試み，情処研報，Vol. 2010-CH-87, No. 8, pp. 1–8 (2010).
[4] 守岡知彦：CHISE による HNG データ収録の試み，漢字字体史研究 二 — 字体と漢字情報（石塚晴通監修，高田智和，馬場 基，横山詔一，編），勉誠出版，pp. 185–203 (2016).
[5] 守岡知彦：CHISE-wiki における HNG カード画像利用の試み，情処研報，Vol. 2016-CH-111, No. 4, pp. 1–8 (2016).
[6] 守岡知彦：データベースの再生と保存についての試論 — HNG を例に—，じんもんこん 2018 論文集，情報処理学会，pp. 373–380 (2018).
[7] 石塚晴通，高田智和，守岡知彦：漢字字体規範史データセット保存会，`http://www.hng-data.org/`.
[8] 守岡知彦，高田智和，石塚晴通，　他：漢字字体規範史データセット，`https://gitlab.hng-data.org/HNG/hng-data/`.
[9] Morioka, T.: Integration of a Chinese character ontology and Historical Glyph Examples, *9th International Conference of Digital Archives and Digital Humanities (DADH 2018)*, Taiwanese Association for Digital Humanities / Dharma Drum Institute of Liberal Arts, pp. 287–300 (2018).
[10] 守岡知彦：Wiki 的手法に基づく構造化データの編集について，人文科学とコンピュータシンポジウム論文集—人文工学の可能性〜異分野融合による「実質化」の方法〜，情報処理学会シンポジウムシリーズ，Vol. 2010, No. 15, 情報処理学会，情報処理学会，pp. 33–40 (2010).
[11] 高田智和：『大字典』データベースをつくる，じんもんこん 2001 論文集，情報処理学会，pp. 221–228 (2001).
[12] 高田智和：大字典データベースをつかう，情処研報，Vol. 2004, No. 58 (2004-CH-62), pp. 45–52 (2004).

© 2019 Information Processing Society of Japan