

漢字字体規範史データセットと単字検索

守岡知彦, 劉冠偉, 高田智和

「漢字字体規範史データベース」は、各時代・各地域（国）には漢字字体の標準が存在し、その標準が変遷することを実証するために作成された「石塚漢字字体資料」（北海道大学石塚晴通名誉教授が在職中に作成、紙カードで蓄積）を前身とし、それを電子化・データベース化したもので、2005 年から 2015 年までの 10 年間、Web 上の検索サービスを提供した。

「漢字字体規範史データベース」の検索サービス停止を受け、利用者がデータセットそのものを長期にわたり安定的に利用できるよう、オープンデータとして、2018 年から「漢字字体規範史データセット」の公開を開始した (<https://gitlab.hng-data.org/HNG/hng-data/>)。「漢字字体規範史データセット」は、全字形用例の画像（紙カード画像）、字形画像（代表字形の切り抜き画像）、字種・用例数などのメタデータから成り、2018 年 9 月現在、漢籍・仏典・国書等の 60 資料を収録している。

「漢字字体規範史データセット」のための検索サービスとしては、「漢字字体規範史データセット単字検索」を開発し、2019 年 5 月からサービスを開始した (<https://search.hng-data.org/>)。従来の「漢字字体規範史データベース」風の検索・表示方法を再現する一方、CHISE-wiki（Character Information Service Environment, <http://www.chise.org/>）上の紙カード画像及び文字情報へのリンクを介した CHISE ベースの検索機能や、Unihan, GlyphWiki など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。また、「漢字字体規範史データセット単字検索」サーバでは、全用例検索、指定した資料内における用例の検索、資料一覧の取得、Web 検索の実行の 4 種の API も提供している。