木簡データベース・木簡庫の公開

1　はじめに―奈文研の木簡に関わるデータベース群

2018年３月、木簡に関する新データベース「木簡庫」を公開した（http://mokkanko.nabunken.go.jp/）。1999年５月の「木簡データベース」の公開以来、約19年ぶりの更新である。本稿では、その経緯と目的、新データベースの特徴などについて述べる。

奈文研では、PC普及以前の1980年代初めから、業務に用いる木簡データベースの構築を開始した。その後、80年代末から90年代初頭にかけての長屋王家木簡・二条大路木簡という厖大な量の木簡出土に対処するためのデータベースの構築を図り、従来のデータベースと統合して1999年、初めての木簡に関するデータベースとして「木簡データベース」を公開した。木簡学会の協力も得て、『木簡研究』に掲載された全国出土の木簡情報も搭載し、日本で唯一の総合的な木簡データベースとして広く活用され今日に至っている。

「木簡データベース」では木簡の全体画像が閲覧できるが、さらに個々の文字を比較検討できるようにしたいと考え、木簡の文字画像データベースを構築し、2005年に「木簡字典」として公開した。その結果、木簡を検索するデータベースと、木簡の文字画像を検索するデータベースの２つのデータベースを併用することになった。

「木簡字典」開発後に取り組んできた課題が２つある。第一は、画像入力によって木簡の文字画像を検索するシステムの開発である。「木簡字典」開発の当初の目的は、木簡解読の参考に資することにあった。しかし、「木簡字典」では文字を特定しなければ文字画像を検索できない。読めていない文字の類似画像の検索は難しいのである。そこで、テキスト入力による文字画像検索システムである「木簡字典」に対し、画像入力による文字画像検索システムを開発した。2016年にPC版、2017年にスマホ・タブレット版を公開したMOJIZO（モジゾー）である（2007年公開の木簡の文字自動認識システムMokkanshop（モッカンショップ）をベースにして開発したもの）。

第二は、「木簡字典」と従来の「木簡データベース」の統合である。「木簡字典」のメタデータは「木簡データベース」のデータに基づいているから、両者のシステムは統合が可能なはずである。ただ、そのためには画像データの大幅な拡充と、個別に順次開発してきた両データベースの長所をいかに活かしながら統合を図るかが、大きな課題であった。

2　新木簡データベース「木簡庫」の構築

新木簡データベースの構築にあたっては、まず既存のデータベースの役割を明確に位置付け直した。すなわち、木簡を検索する「木簡データベース」と、木簡の文字（画像）を検索する「木簡字典」という機能の役割分担を明確化した上で、ユーザインタフェイスの充実を図り、両者を統合することとした。

木簡に関するあらゆる情報資源を収納するクラの意味を込めて、「木簡庫」と命名した新木簡データベースでは、まずトップ画面（図27）において、「木簡をさがす」「文字画像をさがす」の二つの大きなタブの中から、検索方式を選択する構造とした。「木簡をさがす」はさらに「すべて」・「本文」・「カテゴリー」の３つ、「文字画像検索する」はさらに「テキストから」・「画像から」の２つのタブに分け、計５つの機能の中から選択して検索を開始することになる。従来との対応関係は次のようになる。

新 DB「木簡庫」－ 従来の DB 群

「木簡をさがす」

「すべて」検索 －「木簡データベース」の基礎機能

「本文」検索
「項目」検索 } －「木簡データベース」の フィールド指定

「カテゴリー（意味）検索」－「木簡字典」の意味検索

「文字画像をさがす」

文字（テキスト）による検索 －「木簡字典」の基本機能

画像（イメージ）による検索 －「MOJIZO」へのリンク

＊「項目」検索は、「すべて検索」・「本文検索」それぞれに併置。

3　「木簡庫」の特徴と新機能

「木簡庫」の特徴について新機能を中心に整理すると、概ね次の４点にまとめることができる。

a　木簡を検索するデータベースと、木簡の文字（画像）を検索するデータベースの共通プラットフォームの構築

木簡を検索する「木簡データベース」と木簡の文字（画

図27 「木簡庫」のトップページ

像）を検索する「木簡字典」とは開発時期が異なるため、別個に運用されてきたが、相互の関係を明確にし、同一画面から選択して使えるようにした。

b 「木簡をさがす」と「文字画像をさがす」の検索結果一覧の相互移動機能

検索の入口を共通化するだけでなく、「木簡をさがす」の検索結果一覧（テキスト表示）と、「文字画像をさがす」の検索結果一覧（文字画像表示）の相互移動ができ、別個だった２つのデータベースの行き来が可能になった。但し、登録してある木簡すべてに画像がリンクしているわけではないため、画像表示に移動すると、画像が登録してある木簡の文字画像が自動的に選択表示される。

c 多彩な検索機能の効率化と汎用化

「木簡をさがす」では、すべての項目を検索対象とする検索方式をデフォルトとする「木簡データベース」の方針を受け継ぐ一方、検索項目を指定する「木簡データベース」のフィールド検索については、検索頻度の高い木簡本文の検索を「本文検索」として独立させて効率的な検索を可能にするとともに、「木簡字典」で開発した多彩な検索機能を木簡の検索でも使えるようにした。

すなわち、法量についての範囲指定検索（例えば、幅が20㎜から30㎜までの木簡、などの指定）、図にもとづく木簡の形状の検索〈形や切り込みの有無などと型式番号との併用〉、地図を用いた国名検索、タグ付けによる意味検索など、「木簡字典」のみに備わっていた便利な機能を、そのまま「木簡をさがす」に引き継いだ。

d 個別木簡のメタデータからの再検索・リンク機能

「木簡庫」のもっとも深い検索結果である個別木簡のメタデータ一覧（「木簡をさがす」と「木簡の文字をさがす」とで共通）の主な項目について、共通する属性をもつ木簡を直ちに再検索する機能を設けた。また、奈文研の木簡概報を出典とするものについては、奈文研ホームページで公開している「全国遺跡報告総覧」にリンクを張り、当該木簡概報をダウンロードできるようにした。

4 おわりに

「木簡データベース」は奈文研の数あるデータベースの中でもっとも広く利用されてきたものの一つである。これまでほぼ公開当初のまま運用してきたが、改善すべき点が多々生じていた。そこで、木簡の文字画像の切り出しが一定程度蓄積されたのを機に、「木簡字典」の公開以来の懸案だった両データベースの統合を実現して機能の充実を図り、木簡に関する総合的な情報倉庫としての新データベースを構築する運びとなったのである。「木簡庫」と命名した所以である。

研究の推進力としては、東京大学史料編纂所との連携も重要である。「木簡字典」・「MOJIZO」については、東京大学史料編纂所の「電子くずし字字典データベース」との連携検索を実現している。「木簡庫」の開発により「木簡字典」は「木簡庫」の一部として再構築されることになったが、連携検索の根幹はそのまま継承している。一方、「MOJIZO」については、リンクを張るだけで「木簡庫」とは独立した運用を続けることにしたため、「木簡庫」構築の影響は受けていない。

なお、画像による連携検索システムについて、文字のクラ（蔵）に因んで「MOJIZO」という愛称を付したが、「MOJIZO」は語句のニュアンスとしては、画像検索のみでなく、むしろテキスト検索と画像検索を統合したデータベースの呼称としてこそ相応しかったように思う。すなわち、「木簡庫」のうちの「文字画像をさがす」と「電子くずし字字典データベース」との連携全体の呼称に格上げすべきものと思われる。その意味で言えば、従来の「MOJIZO」はいわば「MOJIZOイメージ」、従来の「『木簡画像データベース・木簡字典』『電子くずし字字典データベース』連携検索」は「MOJIZOテキスト」とでも称すべきものということになろう。

データベースは再構築が容易ではなく、何よりも日常的なデータの拡充・メンテナンスが不可欠である。今回約20年ぶりにシステム更新を実現できたが、今後も長期的視野に立って、木簡に関するより使いやすいデータベースのあり方を模索していきたい。

なお、本稿は JSPS科研費JP25220401の研究成果の一部である。

（渡辺晃宏・方 国花）