

文字画像検索システム MOJIZO について

未代誠仁^{†1}

概要: 文字画像検索システム MOJIZO は、「古文書に記された文字の画像をキーとして、デジタルアーカイブ内の文字画像を検索する」ための機能を有する Web ベースの情報システムである。MOJIZO にとって、字形画像を含む古文書デジタルアーカイブの利活用促進は設計当初からの大きな目標である。筆者は、この利活用促進がデジタルアーカイブの構築にかかわった人々、管理者、利用者間のコミュニケーション促進にもつながるものと考えている。MOJIZO には、筆者らが過去に研究・開発を進めていたスタンドアロン型の古代木簡解読支援システム Mokkaishop のために実現した技術と研究活動で得た経験が継承・活用されている。本稿では、Mokkaishop と MOJIZO に関する研究活動のうち、コミュニケーション支援と経験知の共有に関連するものを紹介し、今後の展望について述べる。

キーワード: 字形画像検索, コミュニケーションとデジタルアーカイブ

1. はじめに

古文書デジタルアーカイブの利活用は、歴史学・考古学における重要な課題になっている。デジタル画像として撮影された古文書の点数は日々増加しており、オープンデータ化を含めた公開の動きも盛んである。このような量的な視点においては、利活用は軌道に乗ったと考えることができるが、さらに多方面への利活用を促進していくための議論も必要となっている。具体的には、様々な用途と利用者を見据えた仕組みを目指す施策が必要である。これまで、絞り込み検索機能の拡充、メタデータの共通化による横断検索の実現などの形で進められてきた議論を拡張・進歩させることが求められている。

デジタル画像を主たるコンテンツとする古文書デジタルアーカイブにとって、多方面への利活用に影響する最大の課題はメタデータの整備である。これは、デジタルアーカイブにおける情報検索の機能がメタデータに強く依存しているためである。メタデータの多様化は、情報検索手段の多様化に直結する。したがって、現実的な範囲でメタデータを多様化させる取り組みは不可欠である。

筆者は、古文書デジタルアーカイブにおけるメタデータの多様化とその利活用が、古文書に関わる人々のコミュニケーションを支援する手段の一つになると考えながら、字形画像検索システム MOJIZO の研究、開発を続けてきた[1]。コミュニケーションの多くは、言葉・数値など「何らかの事象を抽象化した」情報のやり取りによって成立する。このとき、求められる抽象化の方法はコミュニケーションの場面ごとに異なる。例えば、物の大きさは数値でもデジタル画像でも表現できるが、それぞれに適した場面というものが存在する。

多くの古文書デジタルアーカイブには、人間が予め高いレベルで抽象化した数値、テキストなどの情報がメタデータとして登録されている。情報検索時には、利用者がメタデータと同レベルの抽象化を行った情報をキーとして入力、

選択することになる。このことは、情報検索における人とコンピュータのコミュニケーションにおいて、抽象度の高い情報がプロトコルになっていることを示している。一方で、デジタル画像のように抽象度が低く事象寄りの情報をプロトコルとする方法の研究はこれからの課題である。デジタルアーカイブは、それ自体が人と機械のコミュニケーションを提供する。さらに、デジタルアーカイブが情報の Hub であることを踏まえると、デジタルアーカイブが扱うことのできる情報の多様化が情報伝達に基づくコミュニケーションの多様化を生み出す可能性は高い。デジタルアーカイブが事象寄りの情報をプロトコルに採用できれば、人と人とのコミュニケーションの多面的にも貢献する可能性は高いといえる。

MOJIZO は、字形画像という比較的事象寄りの情報をキーとして古文書デジタルアーカイブを検索する手段を提供する[2]。字形画像をプロトコルに採用した理由は、筆者が属するグループの研究内容に依るところが大きい。したがって、この単位は絶対のものではなく、古文書デジタルアーカイブの利用方法の整備・拡充に対する取り組みの一つであると筆者は考えている。

本稿では、筆者らがこれまで取り組んできた情報技術の古文書研究への応用に関する研究のうち、コミュニケーションに関連するいくつかを取り上げると共に、今後の展望について述べる。

2. これまでの研究

2.1 Mokkaishop とコミュニケーション

Mokkaishop は、古代木簡解読作業に情報技術を応用する試みとして、スタンドアロン環境における類例字形画像のデジタルアーカイブ検索、解読結果(釈文) / 作業状況の管理・保存などを支援するシステムとして構築したアプリケーションソフトウェアである。Mokkaishop で実現した画像からの字形抽出・解析に関する技術は、後発の MOJIZO にも継承されている。

^{†1} 桜美林大学
J.F.Oberlin University

Mokkanshop では、作業の対象となる古代木簡のデジタル画像に対して利用者が付した釈文、アノテーションなどの情報、適用した画像処理の結果などを記録・管理し、電子メールを介して他の利用者とも共有する機能を有していた[3]。これは、解読作業を行う専門家の知的活動に関する情報を、抽象度に関わらず共有して、専門家同士の共同作業を支援する技術的な取り組みであった(図1)。

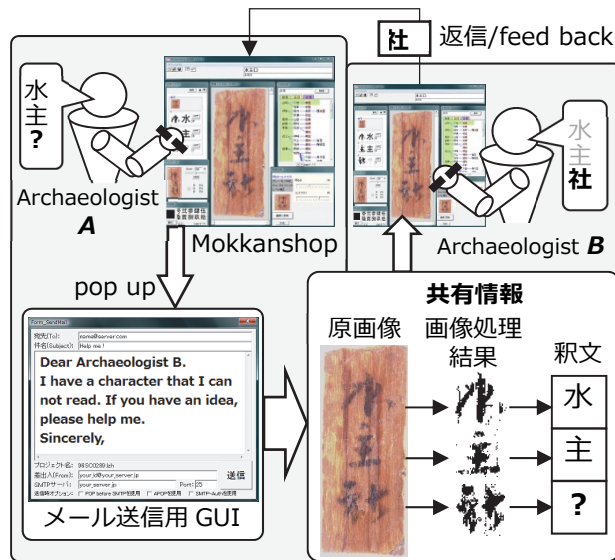


図1 Mokkanshop におけるコミュニケーション
Figure 1 Communication supported by Mokkanshop.

この取り組みでは、情報共有・共同作業のために関係者が Mokkanshop をインストールして使用する必要があった。このため、普及には至らなかったが、研究レベルでの一定の評価を得た。

2.2 字形検索技術とコミュニケーションへの応用

Mokkanshop の普及を妨げた要因の一つは、利用環境にアプリケーションソフトウェアをインストールしないと使えない点にあったと考えられる。この点を踏まえて、字形画像検索サービス MOJIZO ではスタンドアロン環境へのインストールが不要な Web アプリケーションを採用した。MOJIZO の公開により、文字画像をプロトコルとしながら、古文書デジタルアーカイブに記録された経験知を容易かつ広範囲に共有することが可能となった。

MOJIZO は、Web アプリケーションとしてユーザインタフェースを提供する Web サーバと、字形画像を評価して類似字形をリストアップする字形検索サーバの2つのサーバで構成される。このうち、字形検索サーバにはネットワーク経由で字形画像の情報とメタデータを追加・削除する機能を備えている[4]。この機能は当該サーバの字形情報に対する排他的アクセスを必要とし、また Web サーバが管理する情報との整合性を保つ仕組みが未実装のため、MOJIZO では無効化された状態にある。しかし、研究所/研究機関

といった閉じた環境で字形検索サーバを運用できれば、「難読字形+解読結果」のような経験知を効果的に共有するコミュニケーション手段となる可能性を持っている。

3. おわりに

ここまでは、筆者らが取り組んできたコミュニケーションに関連する研究の取り組みを紹介した。最後に、今後の展望を述べて本稿のまとめとする。

経験知には、抽象度の高いものから低いものまでさまざまな情報が含まれる。古文書デジタルアーカイブに関する現在の研究活動は、抽象度が高い情報を優先して取り扱う段階にあると考えられるが、今後は抽象度が低く事象寄りの情報についても利活用を促進する情報技術が提案・実現され、多様な形式を持つ情報が集約・蓄積されていくものと考えられる。

ただし、対面を含む既存の経験知共有手法は今後も継続して利用されるはずである。少なくとも、現在の情報技術が提供可能な情報の表現力は、経験知といわれる情報の多様性に大きく及んでいない。それぞれの経験知が属する世界において、経験知とは何かを分析し、効果的な利活用の方法を模索することは、人文科学・情報学を含む様々な領域において重要な研究分野になると考えられる。そのような研究を支えるのは、当面の間は古来の人的交流にあるのではないかと筆者は考えている。

謝辞

本研究は、科学研究費 基盤(A)-26244041, 基盤(S)-25220401, 基盤(A)-17H00921, 基盤(C)-15K02841 の助成により実施したものである。

参考文献

- [1] MOJIZO: <http://mojizo.nabunken.go.jp/>
- [2] 未代 誠仁, 井上 幸, 高田 祐一, 方 国花, 馬場 基, 渡辺 晃宏, 井上 聡: 木簡およびくずし字のデジタルアーカイブを文字画像で検索するサービスの実装, 情報処理学会 人文科学とコンピュータシンポジウム (じんもんこん 2016) 論文集, Vol.2016, pp.19-24 (2016.12).
- [3] Akihito Kitadai, Yasutaka Tone, Masatoshi Ishikawa, Masaki Nakagawa, Hajime Baba and Akihiro Watanabe: Support System for Standalone and Collaborative Work of Archaeologists to Decode Mokkans, Proc. 12th Conference of the International Graphonomics Society, Melbourne, Australia, Vol.I, pp.226-229 (2007.11).
- [4] 未代誠仁, 白井啓一郎, 馬場基, 渡辺晃宏, 井上聡, 久留島典子, 中川正樹: 古文書字形検索サーバの設計と試作, 日本情報考古学会第 33 回大会講演論文集, Vol.2014(13), pp.75-77 (2014.9).