

## 木簡解読のためのテキスト検索の改善

SHERINI SOMAYEH† 高倉 純† 未代 誠仁‡ 中川 正樹‡ 馬場 基‡‡ 渡邊 晃宏‡‡

†東京農工大学 工学部 ‡東京農工大学 大学院工学府 ‡‡奈良文化財研究所

## 1. まえがき

奈良時代前後の日本では、木片に墨で文字情報を記された「木簡」が文字情報の記録・伝達媒体として広く普及していた(図1)。現在、古代木簡の解読を通して、古代日本に関する様々な研究が進められている。

しかし、発掘された木簡の多くは、1,200年以上前のものである。そのため、汚損や破損によって文字情報の多くが解読困難な状態にある。専門家は、一文字でも多くの文字を解読し、当時の情報を救い上げるために作業を進めている。

木簡解読を行う専門家を支援するため、我々は解読結果となる積文の補完・検証を支援するテキスト検索処理の研究を進めてきた。短文の記述に適した木簡は多くの場合事務的用途に用いられており、定形的な記述が多くみられる。我々は、独自の文脈処理手法である EAC 法(Extended Aho-Corasick 法)と地名/人名/品名/カバネのデータベースの実現を通して、積文補完・検証を効果的に支援するテキスト検索の実現を目指してきた[1][2]。

本報告では、このテキスト検索に関する最新の研究成果について述べる。なお、報告内容は昨年のじんもんこん 2008 の延長上にあるものである[3]。

## 2. 木簡解読支援とテキスト検索

木簡の汚損・破損などに伴って木簡に不可読な

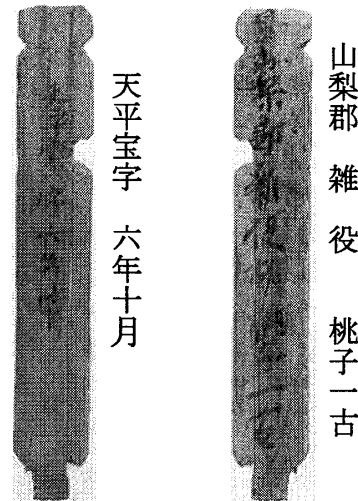


図1 古代木簡の例

部分が発生した場合、専門家は可読部分を他の古文書、同時代について記した書物などと照らし合わせることで、不可読部分の推読に役立つ情報を得ようとする。この手続きを電子化したものが、我々がこれまで提案してきた古代木簡解読支援のためのテキスト検索であった。

テキスト検索の情報化は、専門家を重く分厚い書物群から解放する。また、可読部分をキーとする検索の結果を新たなキーとして、二次的、三次的検索を実行する「芋づる検索」も可能となる。これらは、専門家が解読作業に意識を集中する上で重要な支援となる。一方で、テキスト検索が二次的、三次的に連鎖する芋づる検索では検索空間が爆発的に増加するため、検索履歴の管理機能が必要となる。また、基本的には部分一致検索機能である EAC 法に加えて、可読部分と検索対象の完全一致を条件としたテキスト検索機能も選択できることが望ましい。これらは、テキスト検索のユーザビリティ向上を実現する鍵となる。

Improvement of Text Retrieval to Support Decoding  
Mokkans

Somayeh Sherini†, Jun Takakura†, Akihito Kitadai††,  
Masaki Nakagawa††, Hajime Baba†††, and Akihiro  
Watanabe†††

†Faculty of Engineering, Tokyo University of Agriculture  
and Technology

††Graduate School of Engineering, Tokyo University of  
Agriculture and Technology

†††Nara National Research Institute for Cultural Properties

### 3. 改善版テキスト検索機能

図 2,3 に、今回改善したテキスト検索機能のグラフィカルユーザインタフェース (GUI) を示す。

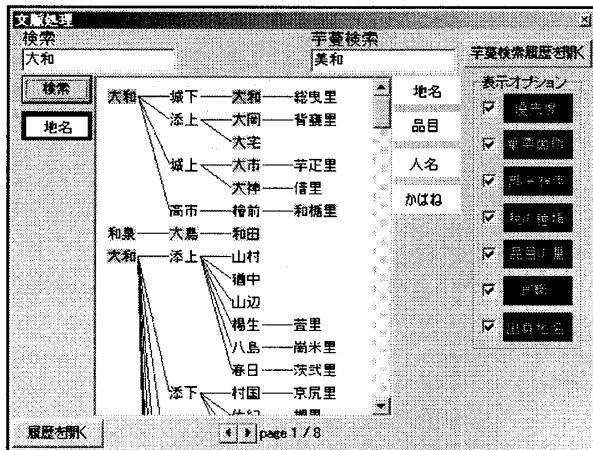


図 2 テキスト検索機能の GUI (履歴表示格納)

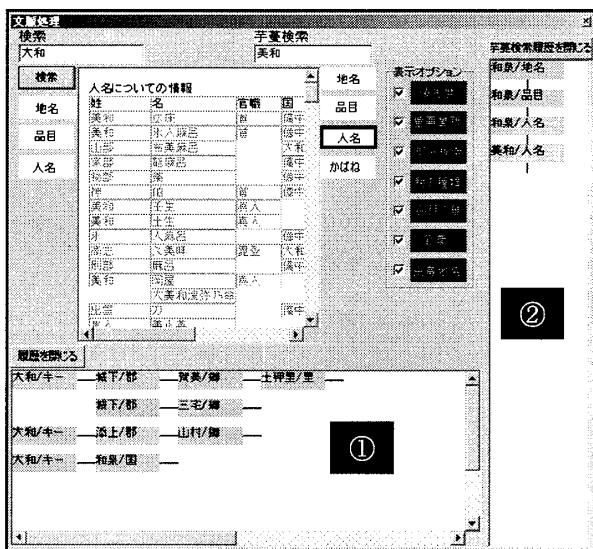


図 3 テキスト検索機能の GUI (履歴表示展開)

図 2 では、検索履歴に関する表示部（下部①および右部②）が閉じられた状態である。一方で、図 3 では履歴表示部が展開されている。これは、画面解像度が小さい Tablet PC などで表示領域を有効に使うための工夫である。

テキスト検索機能が提供するものは、地名/人名/

品名/カバネの各カテゴリを対象とした横断的な検索である。これらのカテゴリのうち、地名だけが国郡郷/国郡里/国郡郷里といったツリー構造を持っている。ツリー構造の情報に対する検索では、同一ツリー内で階層間を行き来するような検索試行が頻発する。このため、他のカテゴリに対する検索履歴を地名に対する検索履歴と同様に扱うと履歴の大部分が地名に対するものとなり、後で履歴を見直す場合、および後述する UnDo/ReDo 機能によって過去の検索試行を復元する場合に妨げとなる危険性がある。そこで、今回は積文を地名の一部と仮定した場合の検索結果を検索履歴部①に、それ以外の検索結果を検索履歴部②にそれぞれ分けて管理・表示した。

履歴を用いた UnDo/ReDo を行う場合は、履歴表示部の該当箇所を電子ペンまたはマウスでクリックする。

### 謝辞

本研究は科研費基盤 S-20222002 および若手 B-19720202 の助成を受けたものである。

### 参考文献

[1] A. V. Aho and M. J. Corasick, "Efficient String Matching: an Aid to Bibliographic Search," Communications of the ACM, vol.18, No.6, pp.333-340, June 1975.

[2] 未代誠仁, 西嶋佳津, 齋藤恵, 石川正敏, 中川正樹, 馬場基, 渡辺晃宏: 木簡解読支援のための文脈処理, 日本情報考古学会論文誌, Vol.13, No.1, pp.7-21 (2007.9).

[3] 高倉純, SomayahSherini, 未代誠仁, 石川正敏, 中川正樹, 馬場基, 渡辺晃宏, "木簡解読支援のための情報検索," 人文科学とコンピュータシンポジウム, IPSJ Symposium Series Vol. 2008, No. 15, pp. 75-80, 2008.