

I. 何謂木簡・草體字解讀系統—MOJIZO—

II. 使用方法

III. 凡例

IV. 圖像準備指南

V. 常見問題

VI. 有關製作人員、相關項目、諮詢等

I. 何謂木簡・草體字解讀系統—MOJIZO—

《木簡・草體字解讀系統—MOJIZO—》*是奈良文化財研究所與東京大學史料編纂所共同開發的，以圖象進行檢索的，“圖象搜索”文字圖象數據庫。“MOJIZO”的漢字為文字藏，具有文字（圖象）庫（藏）的意義。

本資料庫，通過解析用戶所要搜索的圖象，從奈良文化財研究所收藏的木簡字形・字體，以及東京大學史料編纂所收集的古文書・古記錄（貴族的日記）・典籍等字形・字體中，抽出類似的文字圖象，按序顯示。即使無法認讀的文字，只要利用圖象檢索，就能從以上兩機關的數據庫中抽出字形類似的文字，MOJIZO可以說是一個劃時代的新系統。

《木簡・草體字解讀系統—MOJIZO—》為文字圖象資料庫的發展帶來了新的可能性。希望同已往公開的，以文本進行檢索的《木簡圖象數據庫・木簡字典》《電子草體字字典》聯合檢索一同使用。

*MOJIZO：文字藏

英文名：Image matching search for mokkan or cursive characters

II. 使用方法

※有關解析圖象的製作方法，請參看“IV. 圖象準備指南”。



① 解析畫面

起始畫面

[1] 選擇圖象

可以使用以下任意一種方法選擇圖象。

A 直接將文字圖象拖動到 [1] 的位置。

B 點擊 [1] 選擇文字圖象。

C 若是智能手機或平板電腦，可直接點擊 [1] 用手機拍攝要查詢的文字，或者選擇文件夾中的圖象。但是，部分操作系統或機種可能無法正常運行。

[2] 解析

點擊此解析按鈕即可進行解析。解析完畢會顯示②解析結果畫面。

② 結果一覽畫面 1

此畫面分別顯示奈良文化財研究所與東京大學史料編纂所的解析結果一覽。

[3] 顯示用戶選擇的圖象。

[4] 奈良文化財研究所解析結果顯示欄

[5] 東京大學史料編纂所解析結果顯示欄

(以下條目所指內容，[4] [5] 相同)

[6] 類似文字圖象

按照相似度大小，顯示共8件圖象。

要想查看更多圖象，請點擊 [7] 查看更多按鈕。

[7] 查看更多

顯示最多100件圖象。(顯示③畫面。)

[8] 示出圖象所指文字。

[9] 出典顯示欄

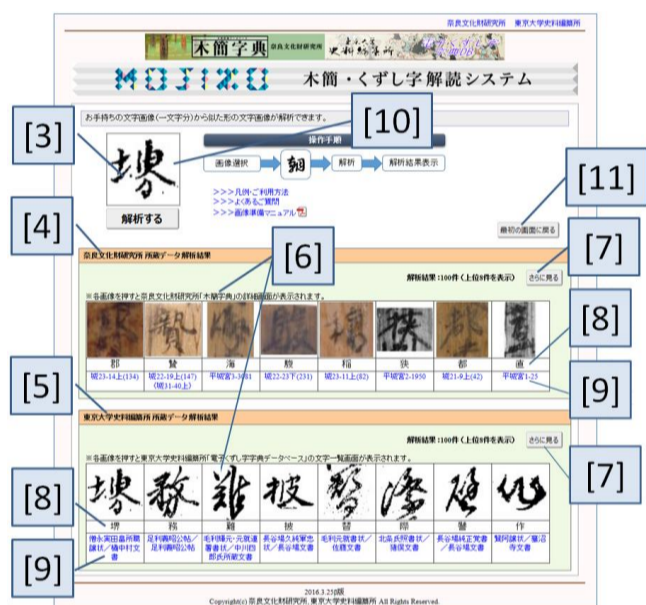
有關 [8] 與 [9] 欄的詳細信息，請參照“III. 凡例”。

點擊各個圖象或出典，將跳轉至奈良文化財研究所的《木簡字典》，或東京大學史料編纂所的《電子草體字字典數據庫》的詳細畫面。

[10] 若想變更所要解析的圖象，請從這裡重新選擇圖象。

[11] 返回首頁

① 返回解析頁面。



③ 結果一覽畫面 2

顯示包括結果一覽畫面1的最多100件圖象，兩個機關的結果各以不同標籤頁顯示。

[12] 用戶選擇的圖象

[13] 文字圖象等(與結果一覽畫面1的構成相同)

[14] 關閉

關閉此畫面。



※智能手機或平板電腦，因屏幕尺寸不同，顯示畫面會有所不同。比如畫面較小的終端設備，解析結果一覽只顯示6件圖象，有些按鈕或解說部分也會被省略。

III. 凡例

1. 此數據庫通過分析用戶所要進行搜索的圖象，從奈良文化財研究所收藏的木簡字形・字體，以及東京大學史料編纂所收集的古文書・古記錄（貴族的日記）・典籍等字形・字體中，抽出類似的文字圖象，分別顯示。但，奈良文化財研究所收藏的資料與東京大學史料編纂所收藏的資料所顯示的結果，其結構有不同之處。詳情請參看以下內容。

2. 需要將圖象修剪成一個文字，因為可以解析的對象只有一個文字。

(有關圖象編輯，請參照“IV. 圖象準備指南”。)



只不過，奈良文化財研究所的解析數據中，包括一部分“戶主”等合字（將多個獨立的字，合寫成一個字，一般由兩個字組成。)

3. 解析結果欄中的文字圖象，按照相似度大小進行排序，與檢索圖象相似度越高，其排列順序越靠前。結果欄畫面，首先顯示前8件圖象，點擊畫面右上方的“查看更多”按鈕，可以閱覽更多圖象，最多為100件。

>>關於畫面的詳細說明，請參照“II. 利用方法”。

4. 解析結果的文字圖象下一欄，只能顯示UTF-8支持的文字。有關超出UTF-8編碼範圍的特殊文字，採取以下方法表示。

【奈良文化財研究所】

用“”表示，然後在()內示出字的結構(例：“”(土+高))。

【東京大學史料編纂所】

以“&#m+大漢和辭典的字元編碼”的形式表示(例：“&#m12646;”)。

5. 解析結果的資料來源詳情如下。

【奈良文化財研究所】

各個出典以正式報告書為優先，如已收錄於正式報告書，將示出其出典(《平城宮木簡》1~7,《平城京木簡》1~3)。

若有訂正，將在()內標明其出典。有關正式報告書中未被收錄的木簡，則記載木簡簡報(《平城宮發掘調查出土木簡概報》4~44)中的相關信息。

例如：“平城宮1-50”=表示《平城宮木簡》1所收錄的50號木簡。

城29-19上(311)=表示《平城宮發掘調查出土木簡概報》29號19頁上段所收錄的編號為(311)號的木簡。

詳情請參考本系統檢索結果欄連結的木簡字典(詳細信息頁面)的出典欄。另外，木簡數據庫也請一併參考。

有關出典的詳細情況，請參照木簡數據庫凡例(<https://www.nabunken.go.jp/Open/mokkan/rei.html>)。

【東京大學史料編纂所】

排列順序為“文書名/史料群名”。

6. 點擊解析結果一覽中的文字圖象，或出典欄，即可顯示該文字的詳細畫面，可以閱覽更詳細的信息。

IV. 圖像準備指南



圖像準備指南 請參照以下方法，編輯手中的需要進行解析的圖象。

可以解析的圖象 — 僅包含一個字
文件大小不得超過 4M
文件格式為 jpg、jpeg、png、gif、bmp
筆跡清晰
背景顏色均勻

準備圖象

(基礎步驟)

- ① 修剪成一個文字
- ② 確認文件大小 (低於 4M)
- ③ 確認文件格式 (僅限於 jpg、jpeg、png、gif、bmp 格式)
請用圖象編輯軟件，選擇適當的保存形式後，重新進行保存，而不是只更改文件擴展名。

古文書

木簡

右側編輯這一部份

請注意！不能包括上下文字的一部分

不能有空白部份

儘量刪除不需要的背景



準備就緒，開始解析圖象

(提高解析精確度)

- ④ 消除背景

橡皮擦工具 或 背景工具 (白色)

儘量消除背景顏色



(進一步提高解析精確度)

- ⑤ 轉為二值圖象 (黑白 2 色)
- ⑥ 調整背景顏色及筆跡

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

橡皮擦工具

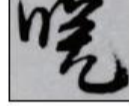
先製作灰度圖象，調節黑白比例之後，再轉換成二值圖象，可以減少背景變黑的部分。

除去背景顏色，或添加筆跡中因破損、污損、蟲損而斷裂的部分。

○ 易于解析的圖象



(木簡)



(古文書)



筆跡清晰、背景顏色均勻，沒有字跡以外的污點。

✗ 難以解析、或不能解析的圖象



包含多字



圖象有空白



背景顏色不均勻



筆跡不清晰



黑底白字



筆跡不清晰



黑底白字



黑底白字

請將圖象顏色進行反轉。

↑ 請參照以上 ① 修剪圖象。

↑ 請參照以上 ④~⑥ 調整背景或筆跡。

➡ 若參照以上方法編輯，也得不到如意的結果，請按照以下方法再試。

自己描繪字的輪廓



線的粗度要接近於實際筆畫的粗度

用鉛筆工具等進行臨摹

或

在白紙上用粗筆摹寫之後再攝影

V. 常見問題

◆ 有關解析圖象

• 解析圖象為何只能是一個文字？

因為奈良文化財研究所與東京大學史料編纂所收錄的數據為一個文字，不對應複數文字或整個資料。基本以一個文字為單位。為了提高解讀的精確度，請將文字圖象修剪成一個文字，再進行解析。但是，奈良文化財研究所的解析數據中包括一部分“戶主”等合字。

• 可以進行解析的圖象有容量限制嗎？

最大容量為4M。如果超出此容量限制，請用圖象編輯軟件調整尺寸後，重新進行保存。若要上傳超出限量的圖象，會提示“無法上傳文件尺寸超過4M的圖象”。

• 有關圖象格式有何要求？

只能解析jpg, jpeg, png, gif, bmp格式的文件。不符合上述格式的文件，即使通過重命名改變其擴展名，也仍然無法正常解析。所以，請用圖象編輯軟件，選擇適當的保存形式，重新進行保存。若要上傳保存形式不適當的圖象，會提示“無法上傳擴展名為jpg, jpeg, png, gif, bmp以外的文件。”。

• 若顯示“無法對此文件進行解析”，該怎麼辦？

以上提示，會在上傳修改文件擴展名的圖象時顯示。請用圖象編輯軟件，選擇適當的保存形式，重新進行保存。

• 可以批量解析嗎？

不可以。請選擇一個文件。若要上傳複數文件，會提示“無法對複數文件進行解析”。

•**點按解析按鈕之後，過了一段時間也不顯示結果。**

如果過了30秒之後也不顯示結果，有可能訪問量過於集中，導致程序不能正常運行。此時，會提示“現在無法進行解析，請稍候片刻再試”。非常抱歉，請將頁面（標籤頁）關閉，稍等片刻，再次進行解析。

•**應該怎麼編輯要進行解析的文字圖象？**

具體請參看“IV.圖象準備指南”。

•**彩色圖象和灰度圖象都可以進行解析嗎？**

本解析系統，不論是彩色圖象還是灰度圖象，都會通過自動處理先將圖象轉換為黑白二值圖象。因為使用圖象的明暗度來進行處理，受到資料的狀態、攝影狀態等影響，識別率會有變低的可能性。請參照“IV.圖象準備指南”，將手中的圖象進行適當的編輯後，再次進行解析。

◆**有關解析結果**

•**為何結果裡不包括正確答案？**

因為此系統僅以形狀為基準，無論文字種類如何，旨在搜索形狀類似的圖象。並且，數據中還包括對於專業人員也棘手的、難解的字形。如果解析結果中沒有如意的答案，請根據“IV.圖象準備指南”編輯圖象後，再次嘗試解析。本系統，現在還不是很完善，為了提高精確度並能為大家提供更加便利的檢索環境，今後我們還將進一步努力。另外，為了提高精確度，我們還在收集並記錄用戶搜索的圖象。

•**MOJIZO是如何對文字圖象進行解析，並找出類似的圖象的？**

將圖象中的黑色部分視為文字，白色部分視為背景，將黑色部分看作筆跡，判斷其形狀後，從數據庫中找出類似的圖象。有關彩色、灰度圖象，利用其明暗度，進行黑白二值化處理。關於不能視為字形的噪點，也會用圖象處理功能儘量除掉。

•**解析結果的順序是怎麼排列的？**

按照相似度高低進行排序，與檢索圖象相似度越高，其排列順序越靠前。

•**解析結果共顯示多少件圖象？**

解析結果一覽畫面中，顯示相似度較高的8件文字圖象。點擊畫面右上方的“查看更多”按鈕，可以閱覽更多圖象，最多為100件。因為按圖象的形狀進行搜索，有一定相似度的搜索結果可以多於100件。然而按相似度排序，100位以後的圖象與所檢索的圖象相似度往往較低，因此本系統將搜索結果限定至100件。另外，解析程序裡含有一部分同一資料的同一文字圖象。因此，結果一覽顯示的文書或木簡的實際圖象，也會有重複的可能性。

※但是，因屏幕尺寸不同，有些智能手機或平板電腦的解析結果一覽畫面只顯示6件圖象。點按“查看更多”按鈕，可閱覽更多圖象。

◆**有關收錄數據及其內容**

•**共收錄什麼樣的文字圖象？**

奈良文化財研究所現階段主要收錄從平城宮・京遺址出土的奈良時代的文字圖象。今後我們還將努力補充飛鳥地區、藤原宮・京遺址出土的木簡，以及各個地區出土的木簡文字圖象。

東京大學史料編纂所收錄前近代，即奈良時代到江戶時代為止的日本史史料（古文書、古記錄、典籍等書寫在紙張上的史料）中的文字圖象。

•**收錄的文字圖象共有多少？**

奈良文化財研究所 5,166件(截至2016年3月)

東京大學史料編纂所 24,296件(截至2016年3月)

◆**關於共同研究**

•**奈良文化財研究所與東京大學史料編纂所為何要共同進行研究開發？**

奈良文化財研究所・東京大學史料編纂所均作為日本史的研究據點，發揮著主導作用。並且，為了將各自的成果公佈於眾，研發了各種資料庫。其中，奈良文化財研究所的木簡圖象數據庫《木簡字典》和東京大學史料編纂所的《電子草體字字典》均收錄各種文字圖象，是為解讀前近代各種文字資料而開發的資料庫，通過網絡得到了大家的廣泛運用。因其基本性格相似，在實際運用時，也可參照雙方的結果，進行對比。預計兩個數據庫的聯合，可以跨越時代，進一步擴展文字圖象的搜索範圍。

詳情請參閱這裡(<http://r-jiten.nabunken.go.jp/setsumei.html>)。

◆**有關再次散佈連結以及版權**

•**若要將圖象用於出版物或學術論文等，需要獲得許可嗎？**

屬於版權法規定的引用，只要是在合理範圍內，不論目的如何，都不需要獲得許可。

超出引用範圍的使用，請聯繫以下各網頁登載的聯絡處。

奈良文化財研究所 研究支援推進部 連攜推進課 文化財情報主管〈若要使用圖象，請查看這裡(<https://www.nabunken.go.jp/fukyu/application.html>)〉

東京大學史料編纂所〈若要使用圖象，請查看這裡(<https://www.hi.u-tokyo.ac.jp/tosho/shiryoriyo.html>)〉

•**想在學校授課時使用，需要獲得許可嗎？**

不用特意獲得許可。

•**可以對本網站進行連結嗎？**

本網站可自由連結。

◆**有關支持環境以及聯繫方式**

•**推薦OS以及瀏覽器都有哪些？**

■OS

Windows Vista、Windows 7、Windows 8、Windows 10、Mac OS X 10.8以上

■瀏覽器

Internet Explorer 11

Google Chrome(最新版)

Mozilla Firefox(最新版)

■需要功能

本系統使用JavaScript（活動腳本）以及style sheet（樣式表）。

請在瀏覽器設置中選擇啟用以上功能之後，再使用。

•**有關係統，如有不明之處，應該向哪裡諮詢？**

對系統及其使用方法如有不明之處，或系統如有不良狀況，請用郵件聯繫相關部門。

•奈良文化財研究所都城發掘調查部史料調查室 [history-labo☆nabunken.go.jp](mailto:history-labo@nabunken.go.jp)

•奈良文化財研究所研究支援推進部總務課 [webinfo☆nabunken.go.jp](mailto:webinfo@nabunken.go.jp)

（請將☆改為@使用。）

•東京大學史料編纂所 >>>跳轉至編纂所的諮詢頁面(<https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html>)

不過，有關個別信息內容，可能給不了您滿意的答案，望請諒解。

VI. 製作人員

- ・奈良文化財研究所
都城發掘調查部史料研究室
渡邊晃宏・馬場基・山本崇・桑田訓也・山本祥隆・井上幸・方國花・藤間溫子
企畫調整部文化財情報研究室
高田祐一
- ・東京大學史料編纂所 電子草體字字典開發項目
久留島典子(代表)・小宮木代良・松澤克行・井上聰・稻田奈津子・遠藤珠紀・山田太造・和田幸大(當時)・宮崎肇
- ・櫻美林大學 総合科學系
末代誠仁

共同研究的經過

奈良文化財研究所與東京大學史料編纂所作為該領域的具有代表性的研究據點，於2009年5月交換了關於數據庫聯合的協議書。並於同年10月公開了《木簡圖象數據庫・木簡字典》《電子草體字字典數據庫》的聯合檢索(<http://r-jiten.nabunken.go.jp/>)。從此以來，通過繼續並強化雙方的聯合協作，作為以下研究項目的一環，受到一起進行共同研究(文字識別系統)的櫻美林大學末代誠仁先生的技術提供，才得以開發此系統。有關聯合的經過，詳情請參閱這裡(<http://r-jiten.nabunken.go.jp/setsume.html>)。

相關項目以及科學研究費補助金

此研究及系統開發為奈良文化財研究所(負責:都城發掘調查部史料研究室)與東京大學史料編纂所(負責:附屬圖象史料解析中心項目電子草體字字典數據庫開發項目(代表 久留島典子))聯合協作的成果。

此外，還包括以下日本學術振興會科學研究費補助金的部分成果。

- ① “為了實現木簡等出土文字資料的資源化而有效利用匯總信息與‘知’的集結功能”
(基礎研究(S), 2013~2017年度(預定), 研究代表:渡邊晃宏, 課題編號25220401)
- ② “以人文・情報學的綜合研究來實現歷史文字的經驗知的資源共用化及多元化分析”
(基礎研究(A), 2014年~2017年度(預定), 研究代表:馬場基, 課題編號26244041)
- ③ “利用信息技術來重新構建歷史知識信息的數據開放及其計劃”
(基礎研究(A), 2014~2018年度(預定), 研究代表:久留島典子, 課題編號26240049)
- ④ “利用信息技術對古文書字形的跨機構的數據存檔進行擴充及利用”
(基礎研究(C), 2015~2017年度(預定), 研究代表:末代誠仁, 課題編號15K02841)

《研究分擔人員》以上各個科研項目的分擔人員如下。()內為所屬部門(當時)。

科研① 研究代表 渡邊晃宏

大山 航(三重大學)
小口 雅史(法政大學)
末代 誠仁(櫻美林大學)
久留島典子(東京大學史料編纂所)
笹原 宏之(早稻田大學)
朱 碧蘭(東京農工大學)
白井 啟一郎(信州大學)
高田 智和(國立國語研究所)
中川 正樹(東京農工大學)
馬場 基(奈良文化財研究所)
森本 晉(奈良文化財研究所)
山口 英男(東京大學史料編纂所)
山本 崇(奈良文化財研究所)

科研② 研究代表 馬場基

井上 聰(東京大學史料編纂所)
末代 誠仁(櫻美林大學)
木村 直樹(長崎大學)
桑田 訓也(奈良文化財研究所)
高田 智和(國立國語研究所)
高田 祐一(奈良文化財研究所)
吉川 聰(奈良文化財研究所)
渡邊 晃宏(奈良文化財研究所)

科研③ 研究代表 久留島典子

有川 正俊(東京大學空間情報科學研究中心)
遠藤 基郎(東京大學史料編纂所)
木村 直樹(長崎大學)
小宮 木代良(東京大學史料編纂所)
近藤 成一(東京大學史料編纂所)
柴山 守(京都大學國際交流推進機構)
馬場 基(奈良文化財研究所)
古瀨 藏(國文學研究資料館)
山家 浩樹(東京大學史料編纂所)
山口 英男(東京大學史料編纂所)
山田 太造(東京大學史料編纂所)

協助方

- ・株式會社DTS WEST(系統開發)
- ・澤村真理子(標誌設計)
- ・方國花(翻譯)

版本說明

2016.3.25 β版

今後仍將隨時進行開發研究，使其功能進一步擴充完善。

公開日: 2016年3月25日

智能手機、平板電腦版公開日: 2017年3月16日

諮詢

對系統及其使用方法如有不明之處，或系統如有不良狀況，請用郵件聯繫相關部門。

(不過，有關個別信息內容，可能給不了您滿意的答案，望請諒解。)

- ・奈良文化財研究所都城發掘調查部史料調查室 history-labo@nabunken.go.jp
- ・奈良文化財研究所研究支援推進部總務課 webinfo@nabunken.go.jp
(請將☆改為@使用。)
- ・東京大學史料編纂所 >>>跳轉至編纂所的諮詢網頁(<https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html>)