

[I. What is "MOJIZO: Image matching search for mokkan or cursive characters"?](#)

[II. How to use](#)

[III. Legend](#)

[IV. Image Preparation Manual](#)

[V. Frequently asked questions](#)

[VI. Personnel in charge of development, Related research projects, Contact information etc.](#)

I. What is "MOJIZO: Image matching search for mokkan or cursive characters"?

"MOJIZO: Image matching search for mokkan or cursive characters" is a database of character images, using images as search items, jointly developed by the Nara National Research Institute for Cultural Properties (hereafter "Nabunken") and the Historiographical Institute of the University of Tokyo (hereafter "HI"). It has been named MOJIZO, implying "a character (moji) storehouse (zo)."

In this database, the search item image is analyzed, and similar images are displayed from both the data accumulated by Nabunken on character forms and styles from mokkan (wooden documents), and similarly from the ancient documents, diaries, and books assembled by HI. This revolutionary system, capable of searching similar character shapes even for illegible characters, has been realized through a linked search. MOJIZO thus opens up new possibilities as a character image database. Please use this alongside the linked search capability, which has already been made publicly available, for characters as text in the Wooden Tablet Image Database (Character Dictionary) and Database of Kuzushi Ji (cursive characters).

II. How to use

Note: Regarding the preparation of images for analysis, please refer to "IV. Image Preparation Manual".



① Search page

This is where the search begins.

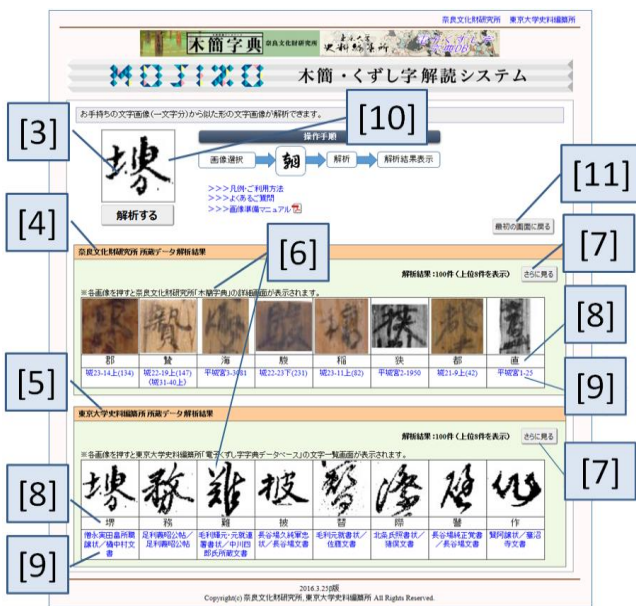
[1] Image selection box

Select an image by either of the following methods.

- A Drag and drop an image file directly into the selection box
- B Click or tap the selection box to browse for an image file.
- C On smartphones, it is possible to tap the selection box and take a photo, which is then directly selected. However, smartphone operation may not work properly depending on the OS or model.

[2] "Analyze" button

Clicking the "Analyze" button starts the analysis. When finished, the Search Results page (see Item ②) will open.



② Search Results page (initial)

This is the initial page displaying the results of the search from the Nabunken and HI databases.

[3] The selected image (search item)

[4] Nabunken results

[5] HI results

(Items given below are common for both [4] and [5])

[6] Similar character images

These are shown from left to right in the order of the degree of similarity.

To display more results, click the following [7] "See more" button.

[7] "See more" button

This will display up to 100 results (the page explained in ③ below will open).

[8] Character corresponding to the image

[9] Character source data

For details on [8] and [9], please see the III. Legend.

Clicking each image or its source data will open the detailed page in Nabunken's Wooden Tablet Image Database or HI's Database of Kuzushi Ji (cursive characters).

[10] To analyze a different character, a new selection can be made here.

[11] Return to the opening page

Clicking this button will display the Search page (①, above).



③ Search Results page (more results)

③ This page opens separately to show more results, up to 100 including the initial items, from either of the two institutes' databases.

[12] Selected image (search item)

[13] Character images, etc. (in the same manner as for the initial results page)

[14] Close button

This will close the current page.

※ For the smartphone and tablet version, because of limited screen size the text and buttons on the screen and the number of results shown in the initial Search Results page (for smartphones it is six results) may differ.

III. Legend

1. In this database a search image is analyzed, and similar images are displayed from both the data accumulated by the Nara National Research Institute for Cultural Properties (hereafter "Nabunken") on character forms and styles from mokkan (wooden documents), and similarly from the ancient documents, diaries, and books assembled at the Historiographical Institute of the University of Tokyo (hereafter "HI"). The configurations differ for the analytic results from data held by the Nabunken and by HI. Please take note of the following content.

2. Search items are single characters. Please prepare the image you wish to be analyzed to have one character only. (Regarding the editing of images, please refer to "IV. Image Preparation Manual".) Analytic data from Nabunken include a portion of ligatures (two characters combined as a single item).

3. The character images in the results are displayed in the order of degree of similarity to the search item. Eight results are displayed in the results window to begin with, and clicking the "See more" button to the right above the images enables viewing up to 100 results.

>>For an explanation of the results window, please see "II. How to use."

4. The characters shown immediately below the image results are limited to items encoded in UTF-8. The following methods are used for indicating special characters not encoded in UTF-8.

(For Nabunken results)

The symbol = will be displayed, and wherever possible an indication given of the composition of the character's form. Example: = (土+鬲)

(For HI results)

The string "&#m" will be given immediately followed by the number for the character in Morohashi's *Dai Kan-Wa Jiten*, plus a semicolon. Example: &#m12646;

5. The data in the field for the image source are as follows.

(For Nabunken results)

For *mokkan* that appear in a formal report, the data for the report (e.g., 平城宮木簡 [Wooden tablets from the Nara Palace Site] Vol. 1-7; 平城京木簡 [Wooden tablets from the Nara Capital Site] Vol. 1-3) are given priority.

Data in angle brackets < > are for revisions of earlier reports. For items not yet included in a formal report, data are given for preliminary reports (平城宮発掘調査出土木簡概報 [Preliminary report on wooden tablets recovered from excavations in the Nara Palace Site] Vol. 4-44).

Examples:

平城宮1-50 = Wooden tablet No. 50 appearing in 平城京木簡, Vol. 1

城29-19上(311) = Wooden tablet 311 (consecutive number) appearing in the upper row, page 19, 平城宮発掘調査出土木簡概報, Vol. 29

For details on each item, please see the field for source on the detailed information page for that item in the Wooden Tablet Image Database (Character Dictionary), which is linked from the search results in this database. See also the data available in the Wooden Tablet Database.

For detailed information on sources, see the Explanatory Notes of the Wooden Tablet Database (in Japanese. <https://www.nabunken.go.jp/Open/mokkan/rei.html>).

(For HI results)

Source data are given as document title/archive (or collection) title.

6. Clicking on an image in the results or on its source data opens a separate page of details from the respective database, where further information can be seen.

IV. Image Preparation Manual

木簡・くずし字解読システム
2016.3.25 β edition

Image Preparation Manual Please refer to the following specifications in preparing the image you wish to analyze.

Images suitable for analysis: Consist of a single character
Are under 4 MB in file size
Are in jpg, jpeg, png, gif, or bmp format only
Are written in clear handwriting
Have a uniform background

Steps in preparation (Please adhere to strictly)

- Trim to a single character (up to 4 MB)
- Check the file size (up to 4 MB)
- Check the file format (jpg, jpeg, png, gif, or bmp only)
Please use image editing software to change to one of the above formats; do not simply change the file extension.
- Erase the background (Use an erasing tool, or a pen or brush tool (to whiten). Leave as little colored background as possible)
- Change color depth to black and white
- Touch up background, character form (Erase the background, and supplement portions of the character lost to breakage, soiling, or insect damage.)

Preparation complete, on to analysis

Examples of images suited for analysis

The handwriting is easily read, the background color uniform, and there are no black portions (from being soiled, etc.) apart from the ink lines.

Examples of images difficult or unsuited for analysis

- Not a single character.
- Has a blank margin.
- The background color is uneven or has gaps.
- The writing is hard to discern.
- The wood grain, soiling, or insect damage are conspicuous.
- There are gaps in the character form.
- The writing is outlined.

Refer to Step ① above and trim examples like these.

Refer to Step ④~⑥ above and touch up the writing or background.

If an acceptable result cannot be obtained even after following the above steps, please try the following method.

Draw the character's outline by hand.
Either copy the form with a pen tool etc.,
or
make an image from a drawing on white paper using a felt pen.
Make the width of the strokes as close as possible to the original.

V. Frequently asked questions

◆Regarding images for analysis

•Why are images for analysis limited to a single character?

As the data held by Nabunken and HI are single-character images, multiple-character strings or entire documents cannot be handled. The basic unit of analysis is a single character. In order to increase deciphering accuracy, please edit your image to a single character and conduct the analysis. The Nabunken data include a number of ligatures (two characters combined as a single item), however.

•What is the upper limit (in MB) on images for analysis?

The maximum is 4 MB. For images larger than this, please use graphic editing software to scale the image down and save the result. Images above the maximum will trigger the error message "Cannot upload files larger than 4 MB."

•Can the image for analysis be in any file format?

Only files saved in jpg, jpeg, png, gif, or bmp format can be analyzed. Simply changing the file extension will not enable analysis; use graphic editing software to save the image in one of the acceptable formats. Incompatible formats will trigger the error message "Cannot upload files other than jpg, jpeg, png, gif, or bmp."

•The message "Cannot analyze this file" was displayed.

This occurs when the file extension only has been changed, but the actual format is unsuitable. Use graphic editing software to save the image in one of the acceptable formats.

•Is it possible to analyze multiple files in batch mode?

This cannot be done. Please select only a single file. Multiple files will trigger the error message "Cannot analyze multiple files in batch mode."

•No results are shown even after clicking the "Analyze" button, even though some time has passed.

If no results show even after 30 seconds or more have passed, the program is possibly not functioning properly due to access overload, etc. In such cases the following error message will be shown: "Not possible to analyze at present. Please wait awhile and retry." Please close the browser or the current tab, and initiate the analysis once again after a brief pause.

•Please tell me how to edit a character image for analysis.

Please consult "IV. Image Preparation Manual" for detailed instructions.

•Can color or grayscale images be analyzed?

In the analysis, color or grayscale images will be converted to black and white. Because this process depends on the image brightness, the conditions of the sample or the lighting when photographed may lower the accuracy of recognition. We recommend that you edit the image according to "IV. Image Preparation Manual" before conducting analysis. Incidentally, submitted images will be stored for later analysis in order to increase the system's accuracy over time.

◆Regarding the analytic results

•The correct character is not among the results.

The purpose of this project is to enable searching similar images regardless of the character type. Also, character shapes being handled may be difficult for researchers to interpret. If the expected character does not appear, please try first to edit the image following "IV. Image Preparation Manual". As this system is still under development, we are continuing our efforts to improve the accuracy and ease of use.

•How is the analysis of similarly shaped characters conducted?

The black portions of the search image are taken as the character and white portions as background, and the shape of the black portions as the character's written form; characters from the databases most closely resembling that form are taken and displayed. Color or grayscale images are converted to black and white based on brightness and then analyzed. Noise separate from the character form is eliminated as much as possible.

•In what order are the results shown?

Images are shown by degree of similarity to the search image, starting from the highest.

•How many search results are shown?

Eight results are shown in the initial Search Results page. By clicking the "See more" button, up to 100 results can be displayed. Because the system relies on the degree of similarity, the number of matches may be very large, but as the degree of similarity beyond 100 images becomes rather low, no more than 100 are shown. Incidentally, as the program sometimes contains multiple images of the same incidence of a character in a particular document, there may be some duplication of individual characters among the *mokkan* or ancient document images shown as the search results.

※For the smartphone and tablet version, because of limited screen size there may be only six results shown in the initial Search Results page. More results can be seen by clicking on the "See more" button.

◆Regarding the content of the recorded data

•What kinds of character images are in the databases?

For Nabunken, the database images are mainly from *mokkan* of the Nara period recovered from the Nara palace and capital sites. The addition of *mokkan* images from the Asuka region, the Fujiwara palace and capital sites, and from *mokkan* recovered at various sites throughout the country is planned for the future.

For HI, the database contains images from premodern Japanese historical documentary materials (historical records consisting of ancient documents, diaries, and books).

•How many character images are included in the databases?

Nabunken: 5,155 items (as of March 2016)

HI: 24,296 items (as of March 2016)

◆Regarding collaborative research

•Why are Nabunken and HI collaborating in the development of this system?

Both Nabunken and HI are core institutions of Japanese historical research, which have fulfilled roles of leadership until the present. Moreover, in order to promote public access and utilization of their research results, they have made public a variety of databases. In particular, Nabunken's Wooden Tablet Image Database (Character Dictionary) and HI's Database of *Kuzushi Ji* (cursive characters) are both databases that have assembled images of the shapes of characters as aids to deciphering a variety of premodern documentary materials, and are offered via the Internet for widespread use by members of society. There are thus many points of similarity in the basic makeup of these databases, and it is assumed they are often utilized in tandem. By linking the two databases, it is anticipated that the scope of character image searches will widen greatly in terms of historic periods.

Please check [here](http://r-jiten.nabunken.go.jp/setsumei.html) for more details(<http://r-jiten.nabunken.go.jp/setsumei.html>).

◆Regarding copyright and linking permissions

•I would like to use the images in academic papers or other publications. Is it necessary to apply for permission?

There is no need to obtain permission for any use that is compatible with fair practice, regardless of purpose, as defined by the Copyright Act. For use going beyond the limits of fair practice, please address inquiries to the respective institutions as follows.

Nabunken: Cultural Property Information Section, Collaboration Promotion Division, Department of Research Support and Promotion (A guide to the use of images is available [here](https://www.nabunken.go.jp/fukyu/application.html)(<https://www.nabunken.go.jp/fukyu/application.html>))

HI: Historiographical Institute, University of Tokyo (A guide to the use of images is available [here](https://www.hi.u-tokyo.ac.jp/tosho/shiryoriyo.html)(<https://www.hi.u-tokyo.ac.jp/tosho/shiryoriyo.html>))

•I would like to use these materials for teaching at school. Is permission required?

There is no special need for permission.

•Is linking to the site permitted?

Please feel free to link to this site.

◆Regarding supported environments and inquiries

Please tell me which operating systems and browsers are supported.

■ OS

Windows Vista、Windows 7、Windows 8、Windows 10、Mac OS X 10.8 or later

■ Browsers

Internet Explorer 11

Google Chrome (latest version)

Mozilla Firefox (latest version)

■ Required functions

This system uses JavaScript and style sheets. Please set your browser to enable these functions.

•Where can I address inquiries about the system?

Please address queries about how to use the system, or notifications of problems, etc., to the following. (Please understand that depending on content, we may or may not be able to respond adequately.)

•Historical Document Section, Department of Imperial Palace Sites Investigations, Nara National Research Institute for Cultural Properties: [history-labo ☆ nabunken.go.jp](mailto:history-labo@nabunken.go.jp)

•Administrative Division, Research Support and Promotion, Nara National Research Institute for Cultural Properties: [webinfo ☆ nabunken.go.jp](mailto:webinfo@nabunken.go.jp)

(replace the ☆ with the @ sign)

•Historiographical Institute of the University of Tokyo >>> [links to HI contact form \(https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html\)](https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html)

Please understand that we may not be able to address all queries regarding particular contents of the data.

VI. Personnel in charge of development

•Nabunken

Historical Document Section, Department of Imperial Palace Sites Investigations

WATANABE Akihiro, BABA Hajime, YAMAMOTO Takashi, KUWATA Kuniya, YAMAMOTO Yoshitaka, INOUE Miyuki, FANG Guohua, FUJIMA Atsuko

Data and information Section, Department of Planning and Coordination

TAKATA Yuichi

•Historiographical Institute of the University of Tokyo, Database of *Kuzushi Ji* Development Project

KURUSHIMA Noriko (project leader), KOMIYA Kiyora, MATSUZAWA Yoshiyuki, INOUE Satoshi, INADA Natsuko, ENDO Tamaki, YAMADA Taizo, WADA Yukio(at the time), MIYAZAKI Hajime

•J. F. Oberlin University

KITADAI Akihito

Collaborative research chronology

As key research institutes representative of Japan in their respective areas, Nabunken and HI signed a joint memorandum in May 2009 regarding database collaboration. Later, this led to making linked search capability in the Wooden Tablet Image and *Kuzushi Ji* databases publicly available(<http://r-jiten.nabunken.go.jp/>) in October of that year (linked search available(<http://r-jiten.nabunken.go.jp/setsume.html>)). Since then, this collaborative relationship has been strengthened, as a component of the research project noted below, with the technical cooperation of J. F. Obirin University's Prof. Kitadai Akihito, who had been pursuing joint research with both institutes from before (in the area of character recognition), the current system was developed.

Related research projects and Grants-in-Aid

The research and development of this system was jointly conducted by Nabunken (handled by the Historical Document Section, Department of Imperial Palace Sites Investigations) and HI (handled by the Database of *Kuzushi Ji* Development Project, led by KURUSHIMA Noriko).

Also, the system is partly the result of research projects supported by the following Grants-in-Aid from the Japan Society for the Promotion of Science.

- ① "Collection and regulation of functional data for converting excavated documentary materials such as *mokkan* into an academic resource"
Grant-in-Aid for Scientific Research (S), 2013–2017 fiscal years (projected), Principal Investigator WATANABE Akihiro, Project No. 25220401
- ② "Combined humanities and informatics research for sharing and converting of knowledge and experience regarding characters in historic materials into an academic resource and for their pluralistic analysis"
Grant-in-Aid for Scientific Research (A), 2014–2017 fiscal years (projected), Principal Investigator BABA Hajime, Project No. 26244041
- ③ "Mechanisms for providing open access to historical knowledge and information, and the restructuring of method to permit effective use of historical information"
Grant-in-Aid for Scientific Research (A), 2014–2018 fiscal years (projected), Principal Investigator KURUSHIMA Noriko, Project No. 26240049
- ④ "Information technology to support the expansion and utilization of a cross-institutional digital archive of character shapes from ancient documents"
Grant-in-Aid for Scientific Research (C), 2014–2018 fiscal years (projected), Principal Investigator KITADAI Akihito, Project No. 15K02841

List of investigators For each of the projects named above, the investigators are given below. Institutional affiliations (at the time) are given in parentheses.

Project ① Principal investigator: WATANABE Akihiro (Nabunken)

OYAMA Wataru (Mie University)

OGUCHI Masashi (Hosei University)

KITADAI Akihito (J. F. Oberlin University)

KURUSHIMA Noriko (HI)

SASAHARA Hiroyuki (Waseda University)

ZHU Bilan (Tokyo University of Agriculture and Technology)

SHIRAI Keiichiro (Shinshu University)

TAKADA Tomokazu (National Institute for Japanese Language and Linguistics)

NAKAGAWA Masaki (Tokyo University of Agriculture and Technology)

BABA Hajime (Nabunken)

MORIMOTO Susumu (Nabunken)

YAMAGUCHI Hideo (HI)

YAMAMOTO Takashi (Nabunken)

Project ② Principal investigator: BABA Hajime

INOUE Satoshi (HI)

KITADAI Akihito (J. F. Oberlin University)

KIMURA Naoki (Nagasaki University)

KUWATA Kuniya (Nabunken)

TAKADA Tomokazu (National Institute for Japanese Language and Linguistics)

TAKATA Yuichi (Nabunken)

YOSIKAWA Satoshi (Nabunken)

WATANABE Akihiro (Nabunken)

Project ③ Principal investigator: KURUSHIMA Noriko

ARIKAWA Masatoshi (Center for Spatial Information Science, University of Tokyo)

ENDO Motoo (HI)

KIMURA Naoki (Nagasaki University)

KOMIYA Kiyora (HI)

KONDO Shigekazu (HI)

SHIBAYAMA Mamoru (Kyoto University International Center)

BABA Hajime (Nabunken)

FURUSE Osamu (National Institute of Japanese Literature)

YANBE Kouki (HI)

YAMAGUCHI Hideo (HI)

YAMADA Taizo (HI)

Cooperation

•DTS WEST Co. (system development)

•SAWAMURA Mariko (logo design)

•Walter Edwards(translator)

Version information

2016.3.25 β edition

Further functional expansion is aimed for through future development and research.

Start of service: 25 March 2016 .

Start for smartphone and tablet version: 16 March 2017

Contact information

Please address queries about how to use the system, or notifications of problems, etc., to the following. (Please understand that depending on content, we may or may not be able to respond adequately.)

•Historical Document Section, Department of Imperial Palace Sites Investigations, Nara National Research Institute for Cultural Properties: history-labo
☆nabunken.go.jp

•Administrative Division, Research Support and Promotion, Nara National Research Institute for Cultural Properties: webinfo☆nabunken.go.jp
(replace the ☆ with the @ sign)

•Historiographical Institute of the University of Tokyo >>>links to HI contact form(<https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html>)