

## I. 何谓木简·草体字解读系统—MOJIZO—

### II. 使用方法

### III. 凡例

### IV. 图像准备指南

### V. 常见问题

### VI. 有关制作人员、相关项目、咨询等

#### I. 何谓木简·草体字解读系统—MOJIZO—

《木简·草体字解读系统—MOJIZO—》\*是奈良文化财研究所与东京大学史料编纂所共同开发的，以图像进行检索的，“图像搜索”文字图像数据库。“MOJIZO”的汉字为文字藏，具有文字（图像）库（藏）的意义。

本数据库，通过解析用户所要搜索的图像，从奈良文化财研究所收藏的木简字形·字体，以及东京大学史料编纂所收集的古文书·古记录（贵族的日记）·典籍等字形·字体中，抽出类似的文字图像，按序显示。即使无法解读的文字，只要利用图像检索，就能从以上两机关的数据库中抽出字形类似的文字，MOJIZO可以说是一个划时代的新系统。

《木简·草体字解读系统—MOJIZO—》为文字图像数据库的发展带来了新的可能性。希望同以往公开的，以文本进行检索的《木简图像数据库·木简字典》《电子草体字字典》联合检索一同使用。

\*MOJIZO：文字藏

英文名：Image matching search for mokkan or cursive characters

#### II. 使用方法

※有关解析图像的制作方法，请参看“IV. 图像准备指南”。



#### ①解析画面

起始画面

##### [1] 选择图像

可以使用以下任意一种方法选择图像。

A 直接将文字图像拖动到 [1] 的位置。

B 点击 [1] 选择文字图像。

C 若是智能手机或平板电脑，可直接点击 [1] 用手机拍摄要查询的文字，或者选择文件夹中的图像。但是，部分操作系统或机种可能无法正常运行。

##### [2] 解析

点击此解析按钮即可进行解析。解析完毕会显示②解析结果画面。

#### ②结果一览画面1

此画面分别显示奈良文化财研究所与东京大学史料编纂所的解析结果一览。

[3] 显示用户选择的图像。

[4] 奈良文化财研究所解析结果显示栏

[5] 东京大学史料编纂所解析结果显示栏

(以下条目所指内容，[4] [5] 相同)

[6] 类似文字图像

按照相似度大小，显示共8件图像。

要想查看更多图像，请点击 [7] 查看更多按钮。

[7] 查看更多

显示最多100件图像。(显示③画面。)

[8] 示出图像所指文字。

[9] 出典显示栏

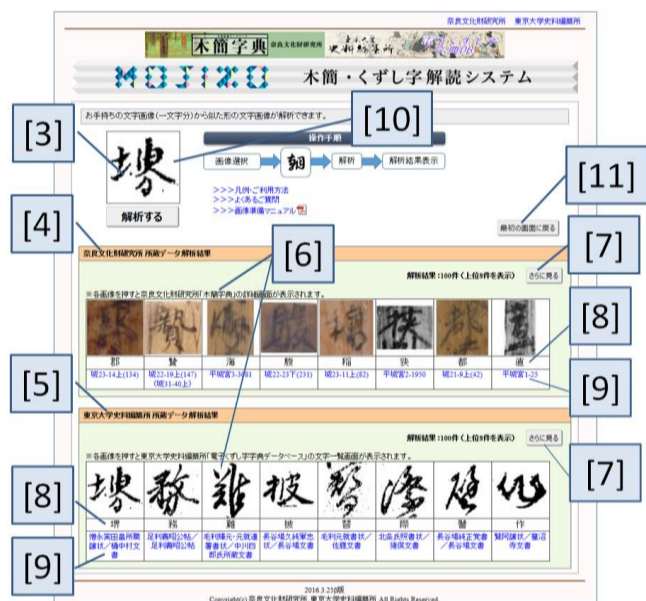
有关 [8] 与 [9] 栏的详细信息，请参照“III. 凡例”。

点击各个图像或出典，将跳转至奈良文化财研究所的《木简字典》，或东京大学史料编纂所的《电子草体字字典数据库》的详细画面。

[10] 若想变更所要解析的图像，请从这里重新选择图像。

[11] 返回首页

①返回解析页面。



#### ③结果一览画面2

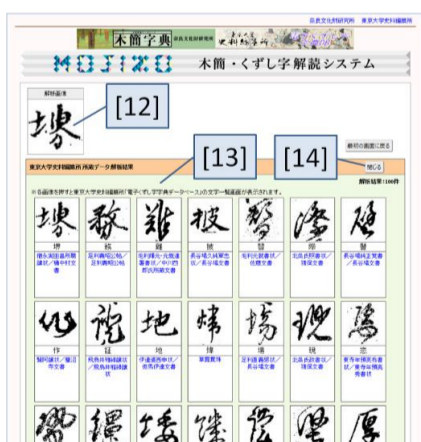
显示包括结果一览画面1的最多100件图像，两个机关的结果各以不同标签页显示。

[12] 用户选择的图像

[13] 文字图像等(与结果一览画面1的构成相同)

[14] 关闭

关闭此画面。



※智能手机或平板电脑，因屏幕尺寸不同，显示画面会有所不同。比如画面较小的终端设备，解析结果一览中只显示6件图像，有些按钮或解说部分也会被省略。

### III. 凡例

1. 此数据库通过分析用户所要进行搜索的图像，从奈良文化财研究所收藏的木简字形·字体，以及东京大学史料编纂所收集的古文书·古记录（贵族的日记）·典籍等字形·字体中，抽出类似的文字图像，分别显示。但，奈良文化财研究所收藏的资料与东京大学史料编纂所收藏的资料所显示的结果，其结构有不同之处。详情请参看以下内容。

2. 需要将图像修剪成一个文字，因为可以解析的对象只有一个文字。

(有关图像编辑，请参照“IV. 图像准备指南”。)



只不过，奈良文化财研究所的解析数据中，包括一部分“户主”等合字（将多个独立的字，合写成一个字，一般由两个字组成。)

3. 解析结果栏中的文字图像，按照相似度大小进行排序，与检索图像相似度越高，其排列顺序越靠前。结果栏画面，首先显示前8位图像，点击画面右上方的“查看更多”按钮，可以浏览更多图像，最多为100件。

>>关于画面的详细说明，请参照“II. 利用方法”。

4. 解析结果的文字图像下一栏，只能显示UTF-8支持的文字。有关超出UTF-8编码范围的特殊文字，采取以下方法表示。

【奈良文化财研究所】

用“”表示，然后在( )内示出字的结构(例：“”(土+高))。

【东京大学史料编纂所】

以“&#m+大汉和辞典的字符编码”的形式表示(例：“&#m12646;”)。

5. 解析结果资料来源详情如下。

【奈良文化财研究所】

各个出典以正式报告书为优先，如已收录于正式报告书，将示出其出典(《平城宫木简》1~7，《平城京木简》1~3)。

若有订正，将在( )内标明其出典。有关正式报告书中未被收录的木简，则记载木简简报(《平城宫发掘调查出土木简概报》4~44)中的相关信息。

例如：“平城宫1-50”=表示《平城宫木简》1所收录的50号木简。

城29-19上(311)=表示《平城宫发掘调查出土木简概报》29号19页上段所收录的编号为(311)号的木简。

详情请参考本系统检索结果栏链接的木简字典(详细信息页面)的出典栏。另外，木简数据库也请一并参考。

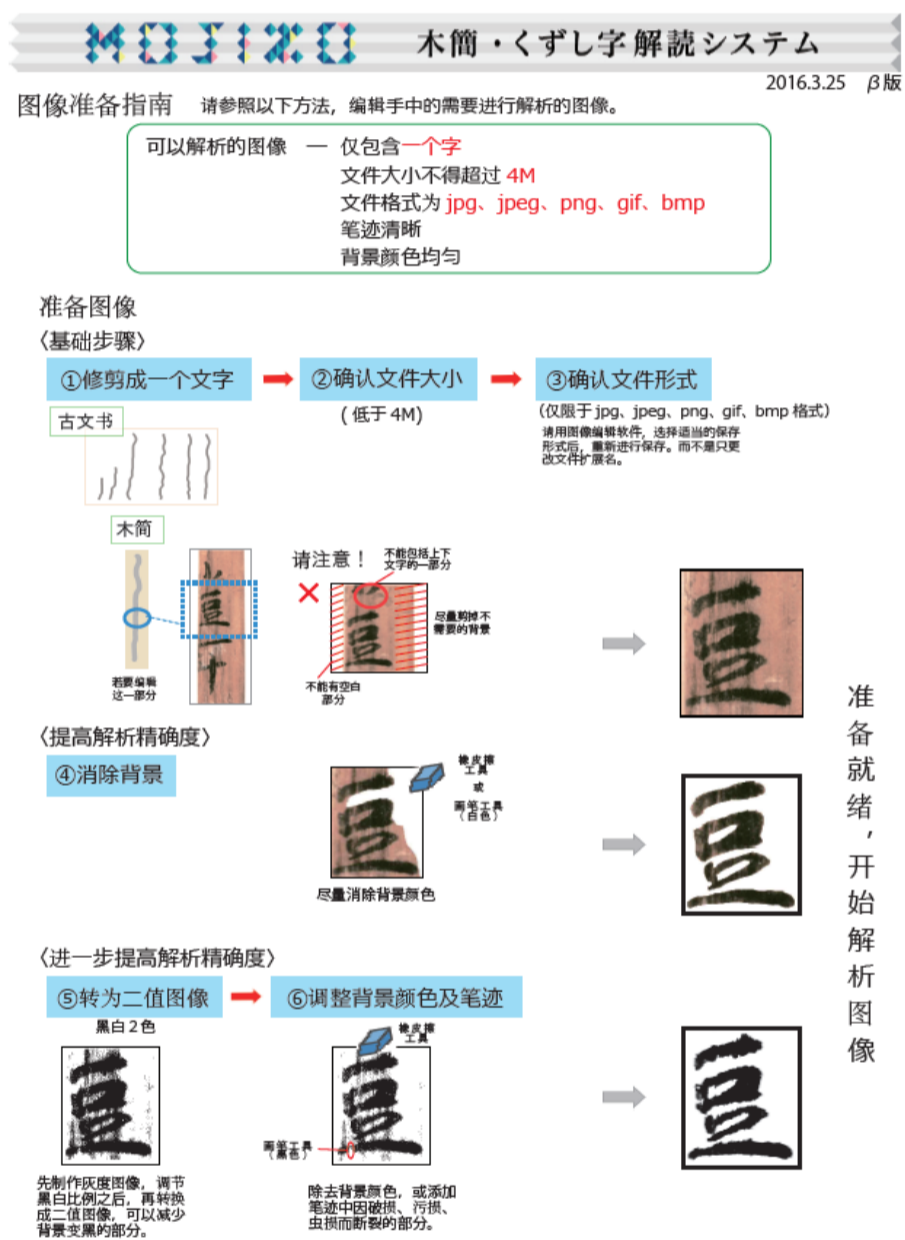
有关出典的详细情况，请参照木简数据库凡例(<https://www.nabunken.go.jp/Open/mokkan/rei.html>)。

【东京大学史料编纂所】

排列顺序为“文书名/史料群名”。

6. 点击解析结果一览中的文字图像，或出典栏，即可显示该文字的详细画面，可以浏览更详细的信息。

#### IV. 图像准备指南



木简・くずし字解説システム 2016.3.25 β版

图像准备指南 请参照以下方法，编辑手中的需要进行解析的图像。

可以解析的图像 — 仅包含一个文字  
文件大小不得超过 4M  
文件格式为 jpg、jpeg、png、gif、bmp  
笔迹清晰  
背景颜色均匀

准备图像 (基础步骤)

①修剪成一个文字 → ②确认文件大小 (低于 4M) → ③确认文件格式 (仅限于 jpg、jpeg、png、gif、bmp 格式)  
请用图像编辑软件，选择适当的保存形式后，重新进行保存。而不是只更改文件名。

④消除背景 (提高解析精确度)  
先制作灰度图像，调节黑白比例之后，再转换成二值图像。可以减少背景变黑的部分。

⑤转为二值图像 → ⑥调整背景颜色及笔迹 (进一步提高解析精确度)  
除去背景颜色，或添加笔迹中因破损、污损、虫损而断裂的部分。

准备就绪，开始解析图像



易于解析的图像

笔迹清晰、背景颜色均匀，没有字迹以外的污点。

难以解析、或不能解析的图像

背景颜色不均匀

笔迹不清晰

黑底白字

请参照以上 ① 修剪图像。

请参照以上 ④~⑥ 调整背景或笔迹。

若参照以上方法编辑，也得不到如意的结果，请按照以下方法再试。

自己描绘字的轮廓

用铅笔工具等进行临摹 或 在白纸上用粗笔摹写之后再摄影

线的粗度要接近于实际笔画的粗度

#### V. 常见问题

##### ◆有关解析图像

##### •解析图像为何只能是一个文字?

因为奈良文化财研究所与东京大学史料编纂所收录的数据为一个文字，不对应复数文字或整个资料。基本以一个文字为单位。为了提高解读的精确度，请将文字图像修剪成一个文字，再进行解析。但是，奈良文化财研究所的解析数据中包括一部分“户主”等合字。

##### •可以进行解析的图像有容量限制吗?

最大容量为4M。如果超出此容量限制，请用图像编辑软件调整尺寸后，重新进行保存。若要上传超出限量的图像，会提示“无法上传文件尺寸超过4MB的图像”。

##### •有关图像格式有何要求?

只能解析jpg, jpeg, png, gif, bmp格式的文件。不符合上述格式的文件，即使通过重命名改变其扩展名，也仍然无法正常解析。所以，请用图像编辑软件，选择适当的保存形式，重新进行保存。若要上传保存形式不适当的图像，会提示“无法上传扩展名为jpg, jpeg, png, gif, bmp以外的文件。”。

##### •若显示“无法对此文件进行解析”，该怎么办?

以上提示，会在上传修改文件扩展名的图像时显示。请用图像编辑软件，选择适当的保存形式，重新进行保存。

##### •可以批量解析吗?

不可以。请选择一个文件。若要上传复数文件，会提示“无法对复数文件进行解析”。

•点按解析按钮之后，过了一段时间也不显示结果。

如果过了30秒之后也不显示结果，有可能访问量过于集中，导致程序不能正常运行。此时，会提示“现在无法进行解析，请稍候片刻再试”。非常抱歉，请将页面（标签页）关闭，稍等片刻，再次进行解析。

•应该怎么编辑要进行解析的文字图像？

具体请参看“IV.图像准备指南”。

•彩色图像和灰度图像都可以进行解析吗？

本解析系统，不论是彩色图像还是灰度图像，都会通过自动处理先将图像转换为黑白二值图像。因为使用图像的明暗度来进行处理，受到资料的状态、摄影状态等影响，识别率会有变低的可能性。请参照“IV.图像准备指南”，将手中的图像进行适当的编辑后，再次进行解析。

#### ◆有关解析结果

•为何结果里不包括正确答案？

因为此系统仅以形状为基准，无论文字种类如何，旨在搜索形状类似的图像。并且，数据中还包括对于专业人员也棘手的、难解的字形。如果解析结果中没有如意的答案，请根据“IV.图像准备指南”编辑图像后，再次尝试解析。本系统，现在还不是很完善，为了提高精确度并能为大家提供更加便利的检索环境，今后我们还将进一步努力。另外，为了提高精确度，我们还在收集并记录用户搜索的图像。

•MOJIZO是如何对文字图像进行解析，并找出类似的图像的？

将图像中的黑色部分视为文字，白色部分视为背景，将黑色部分看作笔迹，判断其形状后，从数据库中找出类似的图像。有关彩色、灰度图像，利用其明暗度，进行黑白二值化处理。关于不能视作为字形的噪点，也会用图像处理功能尽量除掉。

•解析结果的顺序是怎么排列的？

按照相似度高低进行排序，与检索图像相似度越高，其排列顺序越靠前。

•解析结果共显示多少件图像？

解析结果一览画面中，显示相似度较高的8件文字图像。点击画面右上方的“查看更多”按钮，可以浏览更多图像，最多为100件。因为按图像的形状进行搜索，有一定相似度的搜索结果可以多于100件。然而按相似度排序，100位以后的图像与所检索的图像相似度往往较低，因此本系统将搜索结果限定至100件。另外，解析程序里含有一部分同一资料的同一文字的图像。因此，结果一览显示的文书或木简的实际图像，也会有重复的可能性。

※但是，因屏幕尺寸不同，有些智能手机或平板电脑的解析结果一览画面只显示6件图像。点按“查看更多”按钮，可浏览更多图像。

#### ◆有关收录数据及其内容

•共收录什么样的文字图像？

奈良文化财研究所现阶段主要收录从平城宫·京遗址出土的奈良时代的文字图像。今后我们还将努力补充飞鸟地区、藤原宫·京遗址出土的木简，以及各个地区出土的木简文字图像。

东京大学史料编纂所收录前近代，即奈良时代到江户时代为止的日本史史料（古文书、古记录、典籍等书写在纸张上的史料）中的文字图像。

•收录的文字图像共有多少？

奈良文化财研究所 5,166件(截至2016年3月)  
东京大学史料编纂所 24,296件(截至2016年3月)

#### ◆关于共同研究

•奈良文化财研究所与东京大学史料编纂所为何要共同进行研究开发？

奈良文化财研究所·东京大学史料编纂所均作为日本史的研究据点，发挥着主导作用。并且，为了将各自的成果公布于众，研发了各种数据库。其中，奈良文化财研究所的木简图像数据库《木简字典》和东京大学史料编纂所的《电子草体字字典》均收录各种文字图像，是为解读前近代各种文字资料而开发的数据库，通过网络得到了大家的广泛运用。因其基本性格相似，在实际运用时，也可参照双方的结果，进行对比。预计两个数据库的联合，可以跨越时代，进一步扩展文字图像的搜索范围。

详情请参阅这里(<http://r-jiten.nabunken.go.jp/setsume.html>)。

#### ◆有关再次散布外链以及版权

•若要将图像用于出版物或学术论文等，需要获得许可吗？

属于版权法规定的引用，只要是在合理范围内，不论目的如何，都不需要获得许可。

超出引用范围的使用，请联系以下各网页登载的联络处。

奈良文化财研究所 研究支援推进部 连携推进课 文化财情报主管（若要使用图像，请查看这里(<https://www.nabunken.go.jp/fukyu/application.html>)）

东京大学史料编纂所（若要使用图像，请查看这里(<https://www.hi.u-tokyo.ac.jp/tosho/shiryoriyo.html>)）

•想在学校授课时使用，需要获得许可吗？

不用特意获得许可。

•可以对本网站进行链接吗？

本网站可自由链接。

#### ◆有关支持环境以及联系方式

•推荐OS以及浏览器都有哪些？

##### ■OS

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

##### ■浏览器

Internet Explorer 11  
Google Chrome(最新版)  
Mozilla Firefox(最新版)

##### ■需要功能

本系统使用JavaScript（活动脚本）以及style sheet（样式表）。  
请在浏览器设置中选择启用以上功能之后，再使用。

•有关系统，如有不明之处，应该向哪里咨询？

对系统及其使用方法如有不明之处，或系统如有不良状况，请用邮件联系相关部门。

•奈良文化财研究所都城发掘调查部史料调查室 [history-labo@nabunken.go.jp](mailto:history-labo@nabunken.go.jp)

•奈良文化财研究所研究支援推进部总务课 [webinfo@nabunken.go.jp](mailto:webinfo@nabunken.go.jp)

（请将@改为☆使用。）

•东京大学史料编纂所 >>>跳转至编纂所的咨询页面(<https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html>)

不过，有关个别信息内容，可能给不了您满意的答案，望请谅解。

## VI. 制作人员

### ·奈良文化财研究所

都城发掘调查部史料研究室

渡边晃宏·马场基·山本崇·桑田训也·山本祥隆·井上幸·方国花·藤间温子

企画调整部文化财情报研究室

高田祐一

### ·东京大学史料编纂所 电子草体字字典开发项目

久留岛典子(代表)·小宫木代良·松泽克行·井上聪·稻田奈津子·远藤珠纪·山田太造·和田幸大(当时)·宫崎肇

### ·樱美林大学 总合科学系

末代诚仁

## 共同研究的经过

奈良文化财研究所与东京大学史料编纂所作为该领域的具有代表性的研究据点，於2009年5月交换了关于数据库联合的协议书。并於同年10月公开了《木简图像数据库·木简字典》《电子草体字字典数据库》的联合检索(<http://r-jiten.nabunken.go.jp/>)。从此以来，通过继续并强化双方的联合协作，作为以下研究项目的一环，受到一起进行共同研究(文字识别系统)的樱美林大学末代诚仁先生的技术提供，才得以开发此系统。

有关联合的经过，详情请参阅这里(<http://r-jiten.nabunken.go.jp/setsume.html>)。

## 相关项目以及科学研究费补助金

此研究及系统开发为奈良文化财研究所(负责:都城发掘调查部史料研究室)与东京大学史料编纂所(负责:附属图像史料解析中心项目电子草体字字典数据库开发项目(代表 久留岛典子))联合协作的成果。

此外，还包括以下日本学术振兴会科学研究费补助金的部分成果。

①“为了实现木简等出土文字资料的资源化而有效利用汇总信息与‘知’的集结功能”

(基础研究(S)，2013~2017年度(预定)，研究代表:渡边晃宏，课题编号25220401)

②“以人文·情报学的综合研究来实现历史文字的经验知的资源共用化及多元化分析”

(基础研究(A)，2014年~2017年度(预定)，研究代表:马场基，课题编号26244041)

③“利用信息技术来重新构建历史知识信息的数据开放及其计划”

(基础研究(A)，2014~2018年度(预定)，研究代表:久留岛典子，课题编号26240049)

④“利用信息技术对古文书字形的跨机构的数据存档进行扩充及利用”

(基础研究(C)，2015~2017年度(预定)，研究代表:末代诚仁，课题编号15K02841)

《研究分担人员》以上各个科研项目的分担人员如下。( )内为所属部门(当时)。

### 科研① 研究代表 渡边晃宏

大山 航(三重大学)

小口 雅史(法政大学)

末代 诚仁(樱美林大学)

久留岛典子(东京大学史料编纂所)

笹原 宏之(早稻田大学)

朱 碧兰(东京农工大学)

白井 启一郎(信州大学)

高田 智和(国立国语研究所)

中川 正树(东京农工大学)

马场 基(奈良文化财研究所)

森本 晋(奈良文化财研究所)

山口 英男(东京大学史料编纂所)

山本 崇(奈良文化财研究所)

### 科研② 研究代表 马场基

井上 聪(东京大学史料编纂所)

末代 诚仁(樱美林大学)

木村 直树(长崎大学)

桑田 训也(奈良文化财研究所)

高田 智和(国立国语研究所)

高田 祐一(奈良文化财研究所)

吉川 聪(奈良文化财研究所)

渡边 晃宏(奈良文化财研究所)

### 科研③ 研究代表 久留岛典子

有川 正俊(东京大学空间情报科学研究中心)

远藤 基郎(东京大学史料编纂所)

木村 直树(长崎大学)

小宫 木代良(东京大学史料编纂所)

近藤 成一(东京大学史料编纂所)

柴山 守(京都大学国际交流推进机构)

马场 基(奈良文化财研究所)

古濑 藏(国文学研究资料馆)

山家 浩树(东京大学史料编纂所)

山口 英男(东京大学史料编纂所)

山田 太造(东京大学史料编纂所)

## 协助方

株式会社DTS WEST(系统开发)

泽村真理子(标志设计)

方国花(翻译)

## 版本说明

2016.3.25 β版

今后仍将随时进行开发研究，使其功能进一步扩充完善。

公开日:2016年3月25日

智能手机、平板电脑版公开日:2017年3月16日

## 咨询

对系统及其使用方法如有不明之处，或系统如有不良状况，请用邮件联系相关部门。

(不过，有关个别信息内容，可能给不了您满意的答案，望请谅解。)

·奈良文化财研究所都城发掘调查部史料调查室 [history-labo@nabunken.go.jp](mailto:history-labo@nabunken.go.jp)

·奈良文化财研究所研究支援推进部总务课 [webinfo@nabunken.go.jp](mailto:webinfo@nabunken.go.jp)

(请将☆改为@使用。)

·东京大学史料编纂所 >>>跳转至编纂所的咨询网页(<https://www.hi.u-tokyo.ac.jp/footer/inquiry-20160129.html>)。