SAT Submission for the IRG Working Set 2021

Doc. Type: Ideographic Research Group Document
Source: SAT Daizōkyō Text Database Committee (Prof. Masahiro Shimoda)
Status: Member Submission
Action: For consideration by IRG

SAT proposes 383 ideographs for inclusion in the IRG Working Set 2021. The collection consists of 382 SAT characters and 1 SATM character.

SAT characters are collected from Taishō Shinshū Daizōkyō (大正新脩大藏経, 1924; henceforth Taishō). If the shape of a glyph is unclear or dubious on Taishō, additional evidence is supplied from its witnesses, such as woodcut printings and manuscripts. The source names and references of evidence images are listed below:

   - Research Institute of Tripitaka Koreana
   - Private collection of 増上寺 (Zōjōji) Temple
   - Archives of Buddhist Culture, Dongguk University https://kabc.dongguk.edu/
2. 嘉興蔵: 嘉興大蔵経 (Jiaxing Tripitaka). 16th c.-17th c. https://dzkimgs.l.u-tokyo.ac.jp/kkz/
SATM character derives from SAT’s manuscript collection for Buddhist studies. Evidence images are taken from following sources.


The contents of this proposal are found in the following documents:

Appendix 1: Attributes
    IRGN2485_SAT_2021_sub_attributes.xlsx
Appendix 2: TrueType font
    IRGN2485_SAT_2021_sub_glyphs.ttf
Appendix 3: Evidence images
    To be downloaded from https://bauddha.dhii.jp/SAT/ws2021evi.zip
Appendix 4: Auxiliary components
    IRGN2485_SAT_2021_sub_auxcomp_v2.pdf
ISO/IEC JTC 1/SC 2/WG 2/IRG
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITION OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646
Submitters are reminded to:
1. Fill in all the sections below.
2. Read the Principles and Procedures Document (P & P) available at
   for guidelines and details before filling in this form.
3. Use the latest Form from
   https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424SubmissionForm.xlsx
   See also http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424SubmissionForm.xlsx
   for the latest Unifiable Component Variations.

---

**Administrative**

1. **IRG Project Code:** IRG Collection 2021
2. **Title:** SAT Submission for the IRG Working Set 2021
3. **Submitter's Region/Country Name:** SAT Daizókyō Text Database Committee
4. **Submitter Type (National Body/Individual Contribution):** Liaison
5. **Submission Date:** 2021-05-07
6. **Requested Ideograph Type (Unified or Compatibility Ideographs)**
   - If Compatibility, the submitter is strongly encouraged to instead register them as IVS in a new or an existing IVD collection (See UTS #37) with the IRG’s approval (Registration fee will not be charged if authorized by the IRG.).
   - **Unified Ideographs**
   - **Compatibility Ideographs**

7. **Proposal Type (Normal Proposal or Urgently Needed)**
   - **Normal Proposal**
   - **Urgently Needed**

8. **Choose one of the following:**
   - **This is a complete proposal.**
   - **More information will be provided later.**
   - **X**

**B. Technical – General**

1. **Number of ideographs in the proposal:** 383
2. **Glyph format of the proposed ideographs is in TrueType?**
   - **Yes**
3. **Are all the proposed glyphs put into BMP PUA area?**
   - **Yes**
4. **Are data for source references vs. character codes provided?**
   - **Yes**
5. **Source references:**
   - **Do all the proposed ideographs have a unique, proper source reference (member body/international consortium abbreviation followed by no more than 9 alphanumeric characters)?**
   - **Yes**
6. **Evidence:**
   - **Do all the proposed ideographs have a separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?**
   - **Yes**
   - **Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)?**
   - **Yes**
7. **Attribute Data Format: (Excel file or CSV text)**
   - **Excel**
C. Technical - Checklist

### Understanding of the Unification Principles

1. Has the submittor read ISO/IEC 10646 Annex S and does the submittor understand the unification principles?  
   - Yes
2. Has the submittor read the "Unifiable Component Variations" (contact the IRG technical editor through the IRG Convenor for the latest version) and does the submittor understand the unifiable variation examples?  
   - Yes
3. Has the submittor read the IRG PnP document and does the submittor understand the 5% Rule?  
   - Yes

### Character-Glyph Duplication (http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot)

4. Has the submittor checked that the proposed ideographs are **not unifiable** with any of the unified or compatibility ideographs of the latest version of ISO/IEC 10646?  
   - Yes
   - Specify version (e.g. 10646:2012)  
   - 10646:2020
5. Has the submittor checked that the proposed ideographs are **not unifiable** with any of the ideographs in the amendments, if any, of the latest version of ISO/IEC 10646?  
   - N/A  
   - Yes
6. Has the submittor checked that the proposed ideographs are **not unifiable** with any of the ideographs in the proposed amendments, if any, of ISO/IEC 10646?  
   - N/A  
   - Yes
7. Has the submittor checked that the proposed ideographs are **not unifiable** with any of the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Convenor for the latest list)  
   - N/A  
   - Yes
8. Has the submittor checked that the proposed ideographs are **not unifiable** with any of the over-unified or mis-unified ideographs in ISO/IEC 10646? (See Annex E of the IRG PnP document)  
   - Yes
9. Has the submittor checked whether the proposed ideographs have any **similar ideographs** in the current standardized or working sets mentioned above?  
   - Yes
10. Has the submittor checked whether the proposed ideographs have any **variant ideographs** in the current standardized or working sets mentioned above?  
    - Yes

### Attribute Data

11. Do all the proposed ideographs have attribute data including the Kangxi radical code, stroke count, and first stroke (primary)?  
   - Yes
12. Do the proposed ideographs contain secondary radical code and their stroke count and first stroke are also provided?  
   - No
13. Are there any simplified ideographs (ideographs that are based on the policy described in (**Policy Together with the Kangxi radical code** among the proposed ideographs)?  
   - Yes  
   - Does the proposal include proper simplified/traditional indication flag for each proposed ideograph in the attribute data?  
   - Yes
14. Do all the proposed ideographs have the document page number of evidence documents in the attribute data?  
   - Yes
15. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in the attribute data?  
   - Yes
   - If no, how many proposed ideographs do not have the IDS?  
   - Yes
16. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs?  
   - Yes
17. Do all the proposed ideographs contain the total stroke count (kTotalStrokes)?  
   - Yes

---

1 The IRG understands that kTotalStrokes can be ambiguous and subject to different interpretations. The IRG takes no responsibility to check the correctness of the submitted attribute data.
SAT Auxiliary Components for IDS Use (ver. 2)

1. Introduction
This document provides the list of auxiliary components which are and will be found in SAT proposals for IRG Working Sets (WS2017 and thereafter). These components are currently unencoded in UCS, but intended to serve to improve integrity and quality of the IDS-based checking process during deliberation by IRG experts.
The document will be updated as necessary.

2. Nomenclature
Components are named with a convention similar to that of IRGN2225, namely:

\[ SAT - XYYZ \]

Where,
- \( X \) is a Latin letter \( H, S, P, D, \) or \( Z \) which denotes the first stroke type of the component. Each stands for the first stroke type \( 1 (H\text{èng}), 2 (S\text{hù}), 3 (P\text{iē}), 4 (D\text{iān}), \) and \( 5 (Zh\text{è}) \) defined in the IRG PnP Annex K, respectively.
- \( YY \) is a two-digit number denotes the total stroke count of the component.
- \( Z \) is a Latin letter \( A-Z \) that identifies each unique component from those have identical \( XYY \). The letter will be assigned from \( A \) to \( Z \) in the order of our recognition of new component.

The components are represented in IDS with surrounding \& and ;, following the convention of XML character entity reference. For example, a character that contains a component \( SAT - P08A \) would have an IDS such as \( ⿰木&SAT-P08A; \).

3. List
The list of components is ordered according to first stroke, then to the alphanumeric order. Retired entries are painted in gray. Each entry consists of a set of information illustrated below:

<table>
<thead>
<tr>
<th>SAT-XYZ</th>
<th>Preferred IDS decomposition</th>
<th>Ex. of UCS chars using this component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ Other sources</td>
<td></td>
</tr>
</tbody>
</table>

Reference names of external sources:

[CHISE]: [http://www.chise.org/ids/](http://www.chise.org/ids/)

[GW]: [https://glyphwiki.org/](https://glyphwiki.org/)
### 3.1. Horizontal Bar (H)

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT-H06A</td>
<td>大于</td>
<td>Conflated shape related to 车 (U+3694) or 右 (U+2B889).</td>
</tr>
<tr>
<td>SAT-H08A</td>
<td>牙女</td>
<td>Variant of 爨 (U+20A7A) that may occur with or without 工 part.</td>
</tr>
<tr>
<td>SAT-H08B</td>
<td>旨示</td>
<td>Variant of 病 (U+795F) or 奈 (U+67F0), seen on the left side of 款 and 栨.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⇒ CDP-8C6A [CHISE][GW]</td>
</tr>
<tr>
<td>SAT-H12A</td>
<td>日中中木</td>
<td>Variant of 树 (U+6840).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⇒ IRG WS2017 01721 (SAT-06042)</td>
</tr>
<tr>
<td>SAT-H19A</td>
<td>日已日</td>
<td>Variant of 應 (U+5912) or 爨 (U+6182). Similar to 應 (U+2157F).</td>
</tr>
</tbody>
</table>

### 3.2. Vertical Bar (S)

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT-S04A</td>
<td>巴口</td>
<td>Typically an alternative development of 爨 (U+758B) from Small Seal script.</td>
</tr>
</tbody>
</table>
3.3. Slash (P)

SAT-P03A 个又 / 叶又
An alternative development of 支 (U+652F) from the Seal–Cleric glyph.
→ U+301B0

SAT-P06A 閂友 / 閂友
A conflated shape that is related to either 友 (U+53D0) or 天 (U+592D).
The shape without linking between 3rd and 4th stroke still comes under this component.

SAT-P08A 从开
An alternative shape of 矾 (U+22199).
→ U+303E4
3.4. Dot (D)

SAT-D06A 丶平 拴 拴 迨
A cursivized shape of 卒 (U+5352) or 卒 (U+41C2).
The shape with middle dots flipped still comes under this component.

3.5. Turn (Z)

Currently does not have entries.

(End of document)