1. Background

According to WG2 N5119, the report of the Shuowen Seal Ad Hoc group meeting 2019-10-01, the experts formed a consensus to produce the mapping table between the Tonghuaxie version (THX) of Shuowen Jiezi (SWJZ) and Chen Changzhi version (CCZ), which can be a source reference like CJKSrc.txt, EmojiSrc.txt, or TangutSrc.txt. After submitting the full comparison chart between CCZ and the proposed character set N4688 on 2019-10-18, N5133 (2020-04-01) listed the arguable glyph differences. N5187 (2022-06-20) was the response from China and TCA experts; it accepted the glyph distinction by VS instead of the separate encoding for most glyph differences in N5133 if required. Also, there are many modifications of the glyphs of the proposed character set, and they resolved the glyph differences with CCZ. In some cases, the proposed character set keeps the current glyph shape with the differences with CCZ, but no suggestion of their separate encoding. N5190 (2022-02-20) summarizes the glyph changes from the original THX to the latest proposal N5191; some modifications are for the differences listed in N5133, and others are not. Figure 1 summarizes the relationship between N5133 and N5190.

Figure 1: Relationship of the glyph difference and modifications in N5133 and N5190

2. Draft Mapping Table from CCZ to N5191, the Shuowen Seal set based on THX

Considering that N5187 accepted to use of VS to handle the subtle glyph differences in N5133 (especially the type A and B in N5133 categorization), it is possible to draft the major part of the mapping table. This proposal is the draft mapping table for the action item agreed upon in N5119. Some TBD entries exist, but 10454 mapped entries in 11108 are already stable. The stable part would be safe to include for the standard when the standard has the first revision of the code chart of Shuowen Seal.

Figure 2: Preliminary Draft Mapping Table from CCZ to WG2 N5191

① seq_CCZ

The sequential number of the heading Seal glyphs of CCZ, the format is 5 digits with 0 paddings, 00001..11108.
The id assigned to the WG2 N5191 entry corresponds to the id_CCZ in lexical order. There are 2 groups of glyphs: the Seal glyphs that Xu Xuan believed existed before his restructure and the Seal glyphs that Xu Xuan added explicitly (so-called "newly added character," 新附字). The format for the former group is 5 digits with 0-padding, 00001..10706. The syntax for the latter group is X-prefixed 3 digits with 0 paddings, [Xx]001..[Xx]402. The prefix "X" is uppercase in some entries or lowercase in others. Although N5191 does not describe the reason for distinguishing the uppercase prefix "X001" and the lowercase prefix "x301", the values are the same with N5191.

The id assigned to the WG2 N5191 entry, whose glyph has better similarity with the entry of id_N5191_lexical, with the same syntax. There are 2 cases where id_N5191_glyph exists; one is that exactly duplicated glyphs in SWJZ (principle #2 in N5189), and another is that the N5191 glyph is tweaked to avoid the shape conflict with other non-cognate entry (the principle #5 in N5189). For the former case, the id_N5191_lexical is empty because the proposed character set in N5191 removed the exactly duplicated glyph in THX. For the latter case, both id_N5191_lexical and id_N5191_glyphic have the values. The id_N5191_lexical is the id to obtain the cognate character in the N5191; the id_N5191 is the id to find the nearest glyph shape.

The map_type is a classification of the proposed value. There are the following values.

- **Unify** No remarkable glyph difference and should be unified
- **Unify (dup)** No remarkable glyph difference and should be unified with the duplicated entry
- **VS** Visible glyph difference, but N5189 accepted VS to distinguish them.
- **VS (if required)** Subtle glyph difference, but N5189 accepted VS to distinguish them.
- **TBD** No consensus at present
- **Bug_N5191** See the appendix of this document

Figure 3 and Figure 4 illustrate how to determine the map_type.
Figure 4: Selection Process of map_type value for the entries listed in N5190 but not in N5133

③ corresp_UCS (informative)
Corresponding CJK Unified Ideographs to help the review and discussion.

⑥ N5133
The value is the type of glyph difference in N5133. The format is N5133:[A-DGQRa-dgqr]. The letter after “:” is the type of the glyph difference defined in N5133. The uppercase, like N5133:A, is the case in which N5191 resolved the glyph difference between CCZ and the proposed character set. The lowercase, like N5133:a, is the case in which N5189 implicitly accepts using VS to distinguish glyphs (if required).

⑦ N5189
The value is the type of reason for the glyph modification or removal, which N5189 described. The format is N5189:[1-6]. The last digit from 1 to 6 is the type of glyph change defined in N5189. The most important case is “N5189:2”, which indicates that the lexical position of THX is excluded from the proposal because the entry at the position is exactly duplicated (same glyph and same semantics) with other positions. Another important case is “N5189:5”, which indicates the modification of the glyph at the lexical position to avoid the conflict with non-cognate character.

3. Draft Mapping Table with Glyph Images
For easier review, the comparison table with the glyph images is available at
Figure 5 illustrates the table structure with the glyph images and how it shows the attributes. Also, the reduced table

Figure 5: Structure of Mapping Table with Glyph Images
Appendix

The map_type values for seq_CCZ 06469 and 06811 are "Bug_N5191".

The case for 06469 is a wrong duplication of the representative glyph in N5191, as shown in Figure 6. The id_N5191_lexical should be 06249.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Serial No. in Tenghuaxie version</th>
<th>Unicode</th>
<th>TTF Font</th>
<th>Corresp Modern Char.</th>
<th>SW Radical</th>
<th>SW Radical Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>06468</td>
<td>06248</td>
<td>39935</td>
<td></td>
<td>亖</td>
<td>仐</td>
<td>327</td>
</tr>
<tr>
<td>06469</td>
<td>06249</td>
<td>39936</td>
<td></td>
<td>首</td>
<td>仐</td>
<td>328</td>
</tr>
<tr>
<td>06470</td>
<td>06250</td>
<td>39937</td>
<td></td>
<td>稽</td>
<td>仐</td>
<td>328</td>
</tr>
</tbody>
</table>

Figure 6: Bug in N5191 for THX06249

The case for 06812 is a wrong duplication of the serial number of THX in N5191, as shown in Figure 7. The id_N5191_lexical should be 06566, but N5191 set it as 06565.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Serial No. in Tenghuaxie version</th>
<th>Unicode</th>
<th>TTF Font</th>
<th>Corresp Modern Char.</th>
<th>SW Radical</th>
<th>SW Radical Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>06811</td>
<td>06565</td>
<td>39A8C</td>
<td></td>
<td>薬</td>
<td>石</td>
<td>357</td>
</tr>
<tr>
<td>06812</td>
<td>06565</td>
<td>39A8D</td>
<td></td>
<td>薬</td>
<td>石</td>
<td>357</td>
</tr>
<tr>
<td>06813</td>
<td>06567</td>
<td>39A8E</td>
<td></td>
<td>碇</td>
<td>石</td>
<td>357</td>
</tr>
</tbody>
</table>

Figure 7: Bug in N5191 for THX06566.