

Title: Proposal to encode five ideographic tally marks

Source: Ken Lunde (Adobe) & Daisuke MIURA

Status: Individual contribution

Action: For consideration by WG2

References: [L2/15-328](#)

Date: 2016-01-27

Abstract

Tally marks, which form a unary numeral system, have been used for thousands of years as a method of counting or keeping score through the use of stroke clusters. Tally marks also predate modern digits, and in some cases, served as their basis for some scripts. Most tally mark systems represent five digits, which correspond to the number of digits on a single human hand. This proposal introduces an ideographic “tally mark” system, and provides evidence that suggests that it requires a “plain text” representation, and should thus be encoded. Note that the UTC accepted these characters, their tentative names, and their tentative code points on 2016-01-27.

Introduction

Wikipedia’s very brief [Tally marks](#) article describes four tally mark systems, one of which serves as the basis for this proposal.

A tally mark system that is widespread throughout East Asia uses the five strokes of the ideograph 正 (U+6B63) to represent the digits one through five. Once all five strokes are written, which completes the tally cluster, the resulting character looks like the ideograph 正. This system is called 正の字 (*sei-no-ji*) in Japanese, and 正字 (*zhèng zì*) in Chinese. **In this proposal, this system uses character names that begin with IDEOGRAPHIC TALLY MARK.**

Examples

The following tables illustrates how the ideographic tally mark system is used to represent the numbers 1 through 20:

1	2	3	4	5	6	7	8	9	10
一	丁	下	𠄎	正	正一	正丁	正下	正𠄎	正正

11	12	13	14	15	16	17	18	19	20
正正一	正正丁	正正下	正正𠄎	正正正	正正正一	正正正丁	正正正下	正正正𠄎	正正正正

Representative Glyphs

The representative glyphs for the characters in this proposal are used in the tables of this document, reflect the forms that are in common or modern use, and are also provided in the OpenType font that is attached to this PDF file whose glyphs are encoded from the tentative code points in this proposal.

Depending on the font, IDEOGRAPHIC TALLY MARK FIVE may be indistinguishable from CJK Unified Ideograph 正 (U+6B63), but its use as a tally digit may allow it to take on a form that is unique from that ideograph in the same font. *The representative glyph in this proposal uses a sans serif (aka heiti or gothic) design to distinguish it from the corresponding CJK Unified Ideograph in the code charts.*

Tentative Code Points & Proposed Character Names

The characters fit into the existing block named “Counting Rod Numerals” (1D360 . . 1D37F) in the “Notational systems” zone (1D000 . . 1DFFF). The table below provides representative glyphs, tentative code points, proposed character names, and the code points of similar—in shape or meaning—characters:

Glyph	Code Point	Character Name	Similar Characters
一	U+1D372	IDEOGRAPHIC TALLY MARK ONE	U+4E00
二	U+1D373	IDEOGRAPHIC TALLY MARK TWO	U+4E01 U+4E05 U+1D36E
三	U+1D374	IDEOGRAPHIC TALLY MARK THREE	U+4E0B
四	U+1D375	IDEOGRAPHIC TALLY MARK FOUR	
正	U+1D376	IDEOGRAPHIC TALLY MARK FIVE	U+6B63

Properties

The following are the proposed properties for the five characters in this proposal:

```
1D372;IDEOGRAPHIC TALLY MARK ONE;No;0;L;;;1;N;;;;;
1D373;IDEOGRAPHIC TALLY MARK TWO;No;0;L;;;2;N;;;;;
1D374;IDEOGRAPHIC TALLY MARK THREE;No;0;L;;;3;N;;;;;
1D375;IDEOGRAPHIC TALLY MARK FOUR;No;0;L;;;4;N;;;;;
1D376;IDEOGRAPHIC TALLY MARK FIVE;No;0;L;;;5;N;;;;;
```

Emoji Representation

The characters in this proposal are candidates for taking on *emoji style*, such as to use glyphs that appear hand-written, or even animated via the new OpenType ‘SVG’ table.

References, Evidence & Attestation

One of the earliest examples of IDEOGRAPHIC TALLY MARK characters can be found on [page 136](#) of 理論統計学:訂正 (*Riron Tōkeigaku: Teisei*) by 吳文聰 (*KURE Ayatoshi*), published in 1895 by 富山房 (*Fuzanbō*), an excerpt of which is shown below:

計	等	二十三歳	二十二歳	二十一歳	二十歳	十九歳	十八歳	十七歳	十六歳	十五歳以下	
								正正丁		、、、	帝國大學
								正正正		、、、	學習院
								正		、、、	高等商業學校
								正一		、、、	専修學校
								正正正正丁		、、、	專門學校
								下		、、、	慶應義塾大學部
								六三		六三	計

In particular, note how the CJK Unified Ideographs 一 (U+4E00), 丁 (U+4E01), and 下 (U+4E0B) are used to represent IDEOGRAPHIC TALLY MARK ONE, IDEOGRAPHIC TALLY MARK TWO, and IDEOGRAPHIC TALLY MARK THREE, respectively.

The following modern excerpt from a Shizuoka Prefecture (静岡県) [website](#) describes the system for the IDEOGRAPHIC TALLY MARK characters:

「正」をつかったかずのかぞえ方

1	2	3	4	5
↓	↓	↓	↓	↓
一	丁	下	正	正

すきなやさいは、なんですか？

とまと	正 正	➡	とまと	10
にんじん	正 下		にんじん	9
だいこん	正 下		だいこん	8
じゃがいも	正 一		じゃがいも	6



Figure 5.1 on page 105 of a [paper in Japanese](#) contrasts the IDEOGRAPHIC TALLY MARK characters with Western ones:

度数	日本で	外国で
1	一	/
2	二	//
3	三	///
4	正	////
5	正	正
6	正 一	正 /
7	正 二	正 //
8	正 三	正 ///
:	:	:

図 5.1 内外の画線法

A modern English-language [page](#) that teaches how to count in Korean using the IDEOGRAPHIC TALLY MARK characters provides usage examples:

1200 Counting with tally marks

Easy to Learn Korean Words and Phrases *Chad Meyer, Moonjung Kim*

一

One

하나, 일

hana, il

二

Two

둘, 이

dul, i

三

Three

셋, 삼

set, sam

正

Four

넷, 사

net, sa

正

Five

다섯, 오

daseot, o

正 is used for counting things in groups of five.

한자 正(정)은 5개의 그룹으로 물건을 셀때 써요.

hanja jeong-eun daseot-gae-ui geu-rub-euro mulgeon-eul seoltae sseoyo.

I'm going to order the food. Please choose one.

음식을 주문할건데요. 그들 중 하나를 골라보세요.

eumsik-eul jumun-halgeon-deyo. geu-deul jung hana-reul golla-boseyo.

1 vegetable mixed rice, 2 knife-cut noodles, and 3 spicy cold noodles.

비빔밥 1개, 칼국수 2개 와 냉면 3개 시켜주세요.

bibim-bap han-gae, kalguksu du-gae wa naengmyeon se-gae sikyeo-juseyo.

TIP

- 正 is a Chinese character (Hanja) and is used for counting tally marks in the same manner as: I, II, III, IIII and 卅.
- Example: 正正二 is 12 and 正正 is 10.
- 正 is an ideal character for counting as it's composed of five simple horizontal and vertical lines with easy spacing.
- Votes, points, iterations and restaurant menu items can be tallied using this system. These numerals are also used in China and Japan.
- 正 literally means "correct" or "right" and is written as 정 in Korean. 정 (情) can also mean "feelings" and is used to describe a warm-hearted person or close feelings with another person.

비빔밥 : 一

칼국수 : 二

비빔국수 : 三

八

Eight

여덟, 팔

yeodeol, pal

There was [another tally mark system in use in Japan](#) that instead used the five ideographs 一 (U+4E00), 二 (U+4E8C), 三 (U+4E09), 王 (U+738B), and 玉 (U+7389), which formed 玉 though not in the correct stroke sequence, to represent the numerals one through five.

Conclusion

Although ideographic tally marks are often written by hand, which is true of tens of thousands of other characters already in Unicode, the printed evidence presented in this proposal demonstrates their use as genuine characters in running text or in tables, and with properties that indicate their numeric values. This strongly suggests a need to represent them as characters in “plain text” environments. The tally mark system that is in this proposal is unique, and forms a well-established and attested system of counting or keeping score.

Although there appears to be no previous formal proposal to encode tally marks, there was a lengthy discussion on the [Unicode mailing list](#) almost nine years ago, back in February and March of 2007, which began with a post from William J. Poser. Although the discussions were about Western-style tally marks, no particular conclusion was reached.

That is all.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

A. Administrative

1. Title:	Ideographic Tally Marks
2. Requester's name:	<i>Ken Lunde & Daisuke MIURA</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual</i>
4. Submission date:	<i>2016-01-27</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>	
Name of the existing block:	<i>Counting Rod Numerals</i>	
2. Number of characters in proposal:	<i>5</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?		<input checked="" type="checkbox"/>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?		<input checked="" type="checkbox"/>
b. Are the character shapes attached in a legible form suitable for review?		<input checked="" type="checkbox"/>
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Ken Lunde, Adobe Systems Incorporated</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Adobe Systems Incorporated, 345 Park Avenue, San Jose, CA 95110 USA</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?		<input checked="" type="checkbox"/>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?		<input checked="" type="checkbox"/>
7. Special encoding issues:		
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?		<input type="checkbox"/>

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	Yes See text See text
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes See text
4. The context of use for the proposed characters (type of use; common or rare) Reference:	common See text
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes See text
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	Yes Yes See text
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No