

Title: Danda Guidelines Draft
Source: US National Body
Status: For UTC Consideration
Date: 2008-01-30

Rationale

Because no coherent policy on dandas has been explicitly set down in the past, the encoding of dandas has been inconsistent, especially for the scripts in India. The following are guidelines for the encoding of dandas and double dandas (“dandas” henceforth).

Background

The history of encoding the dandas in Unicode goes back to 1991, when a single danda (U+0964 DEVANAGARI DANDA) and double danda (U+0965 DEVANAGARI DOUBLE DANDA) were encoded as part of Unicode 1.0 by virtue of their status in ISCII. The dandas were encoded only once, implicitly to be used in the nine official scripts of India (Devanagari, Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telegu, Kannada, and Malayalam). To date, the dandas are still used in these scripts, so the addition of script-specific dandas for any of these nine scripts would significantly impact existing and legacy data. ISCII 1991 encoded the danda at 0xEA and labeled it “Full Stop (Viram, Northern Scripts),” which implied common usage. In the comparative script charts that showed how ISCII codes were to be interpreted for each script, it showed nearly identical glyphs for the viram of nearly all Northern scripts (with the possible exception of Oriya, which has a slightly taller form). (The southern scripts used full stops, instead of the danda.)

Guidelines

1. For non-Indic scripts:

Dandas occurring in Brahmi-based scripts outside of India will be separately encoded.

2. For Indic scripts:

For the scripts of India, a number of criteria will be used in order to determine whether script-specific dandas may be necessary.

The rationale for not encoding script-specific punctuation is that user communities could be disadvantaged if characters are encoded that are, in most contexts, very similar in shape, so that users must choose from nearly-identical characters. Also, since dandas are punctuation marks, their use across scripts would be consistent with other marks of punctuation.

The key issue is how to determine whether a danda in one script is clearly different from another and hence should be separately encoded, particularly given that fonts and styles may offer a range of variation.

If the danda shape (typically the Devanagari danda) is not acceptable to the normal user in the target language, then separate encoding may be warranted.

The test that should guide proposers is to present to a normal user a series of texts containing a danda from another script (most often the Devanagari danda) that has been created with different fonts and see if the normal user can identify anything wrong. For example, if a text in Gujarati had a danda belonging to a different script, would the user be forced to make a shape change? If so, there is good reason to encoding a script-specific danda.

Other documents:

1: Script Specific Danda and Double Danda, by Eric Muller, 6 Nov. 2004, <http://www.unicode.org/~emuller/iwg/p1/>

Public Review Issue #59, Disunification of Dandas, L2/05-049, 2005

Comments on PRI #59: Script Specific Danda and Double Danda, L2/05-055, by Eric Muller, 29 January 2005