

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation internationale de normalisation
 Международная организация по стандартизации

Doc Type: Working Group Document**Title: Proposal to encode two COMBINING HEART characters in the UCS****Source: Michael Everson, Roozbeh Pournader, and John Cowan****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2003-04-01**

One user group has been overlooked during the development of the UCS, namely that of young adolescent women. In order to support their text-encoding requirements, it is necessary to add two characters, COMBINING HEART ABOVE and COMBINING HEART BELOW, and to provide advice to users as to its implementation in a number of scripts.



COMBINING HEART ABOVE



COMBINING HEART BELOW

The quintessential use of this character is as an ordinary combining ornament. Its use is script-independent:

little like losing interest inside

Latin: *wishing you a happy and joyous birthday*

Arabic:

سحر جونم
 تولدت
 شونصد و پنجاه و یک هزار بار
 مبارک
 النار
 بهار ۸۲

سحر جونم
 تولدت
 شونصد و پنجاه و یک هزار بار
 مبارک
 النار
 بهار ۸۲

Note that when used with the Arabic script (as, for instance, in setting the Persian sample shown here), the COMBINING GRAPHEME JOINER should be used to cluster the COMBINING HEARTS into pairs and triplets, for use with the undotted versions of Arabic letters.

Alternatively, since Middle-Eastern and Indic scripts get their own combining characters which are not shared by Latin, Greek, and Cyrillic, suggest these combining sequences to be encoded for Arabic: ARABIC HEART ABOVE, ARABIC TWO HEARTS ABOVE, ARABIC THREE HEARTS ABOVE, ARABIC HEART BELOW, ARABIC TWO HEARTS BELOW, ARABIC THREE HEARTS BELOW. Additional combinations could be expected evidence is found of their usage in minority Arabic-script languages, and possibly Hebrew and Indic scripts.

This character is important for the self-expression of the primary user group. Graphologists have recognized its use as a marker of personality type. From <http://www.graphology.usallportal.com/f421.html>:

i-dot formed like a heart: Warm, generous and giving, these people are great to have as a friend, but they will seek acknowledgment from you or they may feel rejected. They can be demanding at times for attention, but all in all they have a good intent more times than not.

To the right here is a sample of the character as used in a commercial context, advertising chocolate.



We recognize that WG2 and the UTC may not prefer the character-encoding approach proposed here. In such a case, in order to ensure that the user group’s requirements are met, we propose the following:

- a) A control character approach: a control character that appears before a character and indicates that all of its dots are to be represented as hearts. This can be used with characters like those shown below:

U+05B1		HEBREW POINT HATAF SEGOL
U+0687		ARABIC LETTER TCHEHEH
U+071E		SYRIAC LETTER YUDH HE
U+0FCB		TIBETAN SYMBOL NOR BU GSUM -KHYIL
U+1368		ETHIOPIC PRAGRAPH SEPARATOR
U+0C25		TELUGU LETTER THA

- b) A variation selector approach, a variant selector that changes the dots of the previous character into hearts. This may be preferable to the previous approach because the list of characters to be used with this can be limited by WG2 and the UTC, which will help alleviate security concerns, since otherwise, unscrupulous users may mischievously employ the previous control character before a U+2022 BULLET or U+22C5 DOT OPERATOR and make either of them appear as U+2661 WHITE HEART SUIT.