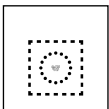ISO/IEC JTC1/SC2/WG2 N2822

L2/04-268

2004-06-20

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 add INVISIBLE LETTER to the UCS**
Source:	**Michael Everson, Peter Constable, Rick McGowan, & Ken Whistler**
Status:	**Individual Contribution**
Date:	**2004-06-20**

Although the Unicode Standard has long recommended the use of SPACE before combining marks in implementation when the user wishes to display such a mark in isolation, implementation practice on a number of platforms has shown this to be problematic, causing difficulties for users. This is not necessarily the *fault* of any particular implementation, and even where applications and systems which *are* capable of displaying SPACE + a combining mark correctly, in XML and HTML applications on those platforms, the behaviour is still compromised because XML and HTML typically collapse sequences of spaces to a single space.

A solution to this is proposed here. A single character, with the LO "letter other" property, the AL "alphabetic" linebreak property, the ON "other neutral" bidirectional property, the NA "Narrow" East Asian width property, and with the Qaai "script inherited" property, can be used as a spacing glyphless letter to which diacritical marks can be attached for display. The typical representation between two words would be <space, invisible-letter, combining-mark(s), space>. The character would *not* be Default Ignorable; if a system does not support displaying it as nothing, the system should display it with something, such as its "nondisplayable glyph" representation. We propose to encode the character at U+2427, in the Control Pictures block, along with similar characters the BLANK SYMBOL and the OPEN BOX, which are used as visible graphics for SPACE.

INVISIBLE LETTER
• can be used as a base letter to display combining characters in isolation

A. Administrative
1. Title
Proposal to add INVISIBLE LETTER to the UCS.
2. Requester's name
Michael Everson, Peter Constable, Rick McGowan, & Ken Whistler
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2004-06-20
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
Proposed name of script
1b. The proposal is for addition of character(s) to an existing block
Yes.
1b. Name of the existing block
Control Pictures.
2. Number of characters in proposal
1
3. Proposed category (see section II, Character Categories)
Category A.
4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)
Level 1.
4b. Is a rationale provided for the choice?
Yes.
4c. If YES, reference
Spacing character.
5a. Is a repertoire including character names provided?
Yes.
5b. If YES, are the names in accordance with the naming guidelines in Annex L of ISO/IEC 10646-1: 2000?
Yes.
5c. Are the character shapes attached in a legible form suitable for review?
Yes.
6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson. TrueType.
6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson. Fontographer.
7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
No.
7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
No.
8. 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)?
Yes.
9. 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/Public/ UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
No.
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
The problems describe above have been discussed by users on public discussion lists.
2b. If YES, with whom?
2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
Yes.
4a. The context of use for the proposed characters (type of use; common or rare)
Common when it is required to display combining characters in isolation.
4b. Reference
5a. Are the proposed characters in current use by the user community?
No.
5b. If YES, where?
6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
General usefullness for most BMP scripts.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
N/A.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
12b. If YES, reference
13a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
13b. If YES, describe in detail (include attachment if necessary)
14a. Does the proposal contain any Ideographic compatibility character(s)?
No.
14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?
14c. If YES, reference