

JTC1/SC2/ WG2 N4310

To: ISO/IEC JTC1/SC2 WG2
 Title: Proposal for Four Characters for Bidi
 From: Aharon Lanin, Mark Davis, and Roozbeh Pournader
 Status: Liaison Contribution (Unicode Consortium)
 Date: 31 July 2012

HTML/CSS recently introduced “bidirectional isolates” to improve handling of bidirectional text in HTML. However, this new technology does not provide a means to solve the bidi issues in non-HTML documents or when copying and pasting HTML into plain text.

This proposal requests four format characters that can be used to support formatting of bidirectional text in non-HTML documents and plain text, in a way which can be interoperable with the mechanisms used by HTML/CSS for markup:

1. LRI U+2066 LEFT-TO-RIGHT ISOLATE

This character, abbreviated “LRI,” marks the beginning of a left-to-right bidirectional isolate.

2. RLI U+2067 RIGHT-TO-LEFT ISOLATE

This character, abbreviated “RLI,” marks the beginning of a right-to-left bidirectional isolate.

3. FSI U+2068 FIRST STRONG ISOLATE

This character, abbreviated “FSI,” marks the beginning of a first-strong bidirectional isolate, i.e. one whose direction is determined by applying algorithmic paragraph rules to the isolate’s content as if it were a separate paragraph.

4. PDI U+2069 POP DIRECTIONAL ISOLATE

This character, abbreviated “PDI,” is used to mark the end of a bidirectional isolate.

Because these characters have been approved by the Unicode Consortium, a liaison member, and will be incorporated into a revised version of the Unicode Bidirectional Algorithm, we consider the characters to be urgently needed, and ask they be added to DAM1.

Character properties

2066;LEFT-TO-RIGHT ISOLATE;Cf;0;LRE;;;;;N;;;;;

2067;RIGHT-TO-LEFT ISOLATE ;Cf;RLE;;;;;N;;;;;

2068;FIRST STRONG ISOLATE ;Cf;LRE;;;;;N;;;;;

2069;POP DIRECTIONAL ISOLATE;Cf;;PDF;;;;;N;;;;;

**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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

A. Administrative

1. Title:	<i>Proposal for Four Characters for Bidi</i>
2. Requester's name:	<i>Aharon Lanin, Mark Davis, and Roozbeh Pournader</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Liaison (from Unicode Consortium)</i>
4. Submission date:	<i>31 July 2012</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):		
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<i>yes</i>	
Name of the existing block:	<i>General Punctuation</i>	
2. Number of characters in proposal:	<i>4</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?		
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Roozbeh Pournader</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Roozbeh Pournader</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Not applicable</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Not applicable</i>	
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)?	<i>no</i>	

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: N4102-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	<i>no</i>
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:	<i>yes</i> <i>Implementers of bidi</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Modern users of bidi</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>(not in use yet)</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>not yet</i>
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:	<i>yes</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>yes</i>
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:	<i>no</i>
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:	<i>no</i>
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:	<i>no</i>
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:	<i>no</i>
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)	<i>Yes</i> <i>See doc.</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>no</i>