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 signs for external and internal links in the UCS
Author: Karl Pentzlin
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2012-05-03
Supersedes: L2/06-268

Block: Miscellaneous Symbols and Arrows

Link Signs

 U+2Bx1   EXTERNAL LINK SIGN

 U+2Bx2   INTERNAL LINK SIGN

I. Use and semantics of the proposed characters

The proposed EXTERNAL LINK SIGN marks external links within web pages (i.e. links which lead to another site, contrary to internal links which lead to another page within the same site or domain).

The proposed INTERNAL LINK SIGN marks internal links within web pages (i.e. links which lead to another page within the same site or domain).

Such a marking on external links especially used where this has legal implications (such as denying responsibility for the contents reached by the marked external link, or claiming rights on the information reached by the internal link), thus acting in a somewhat similar way as U+00A9 COPYRIGHT SIGN, U+00AE REGISTERED SIGN or U+2122 TRADE MARK SIGN (which also were the models for the chosen naming of the proposed symbol).
As the information about legal facts or claims is an important part of the semantic field of the marking and therefore of the proposed symbol, the symbols are not pure control symbols (denoting only the external/internal link at a technical level), unlike some other link symbols commonly found on web pages as "previous" or "home". Therefore, they are not directly comparable in function to e.g. a carriage return or its symbols (U+000D, U+240D, U+2386, U+23CE).

They are used as a part of the descriptive text the user has to click (or consider) to call the link (mostly as the first character of that text, sometime as the last character). Thus, they are used as a textual characters within the plain text which constitutes the link description.

As the symbol occurs only as part of a text (unlike e.g. the aforementioned "previous" or "home" symbols), they are no pure "icons" which carry their meaning by itself (like e.g. a traffic sign). They are an integral part of the text which they denote (the link description), like a TRADE MARK SIGN does not occur without the text which it denotes (the trade mark name).
Thus, they are characters rather than icons.

The designs shown below (see "specimens") all show a design which fits into the text, having a height comparable to the ascent of the default font as long as the default font size in common HTML viewers is used. The symbols are also commonly used within running text. This shows they are used like characters, and thus deserve to be encoded as characters, rather than to be standardized as a graphical icons which cannot be assumed to fit universally into running text.

As these characters occur only in running text at semantically determined positions, they can be regarded as punctuation marks.
Following this, it can be decided that they get the General Category Po instead of So, and are placed in the block Supplemental Punctuation (U+2E00...U+2E7F).

The proposed characters are by no means to be considered as *"the"* external/internal link signs, i.e to be an attempt to standardize symbols for the "external/internal link" functionality on web pages (or to be normative in any other way). Therefore, the addressed institution for standardizing it is not the W3C or a similar institution.
They to be considered as *"an"* external link sign and *"an"* internal link sign which are really found "in the wild" used as characters (as discussed above). Therefore the Unicode consortium is addressed to standardize them. It is like encoding a script character without saying anything normative about the orthographies using that script or telling anybody when or how to use it, yet giving it a name based on its function (e.g. common phonetic value) rather than its shape.

In short: The "external/internal link signs" are proposed to be encoded because they are there, not to standardize any use.

II. Appearance of the proposed characters

In its essence, the EXTERNAL LINK SIGN symbol is a rectangle with an arrow starting from its center, with the arrowhead outside. The rectangle usually has a gap, dent or sloping where the arrow crosses it. The reference glyph above shows a single-stemmed arrow pointing diagonally upward ("north-east"). Glyph variants show a double-stemmed arrow (e.g. at Wikipedia) or an arrow pointing rightward (see specimens).
The proposed reference glyph is designed to be the "common denominator" of the glyphs found. As it applies to Unicode characters generally, the reference glyph is in no way intended to be taken for a standardized icon or to be the best design in any font.
This is also the reason to not name the symbol after the appearance of its reference glyph, e.g. something like "RECTANGLE WITH NORTH WEST POINTING ARROW", as it is appropriate for technical symbols like U+21B5 DOWNWARDS ARROW WITH CORNER LEFTWARDS.

The INTERNAL LINK SIGN is in its essence a rectangle with an arrow pointing into the same direction as shown in the EXTERNAL LINK SIGN, but being confined within the closed rectangle.

As no specimens were found within right-to-left scripts, it only can be speculated (or left to future font designers) whether the arrow would point leftward with such scripts or straight upward within mixed direction environments. This is considered to be within the possible range of glyph variation. It is not proposed to set the Bidi_Mirrored property to "yes", as the proposed symbols are neither paired punctuation nor a math operator (which are the types of characters currently covered by Bidi_Mirrored).

III. Properties of the proposed character

All properties of the proposed character are proposed as the ones common to symbols.
Thus, the entry in UnicodeData.txt is proposed as:

> 2B*x*1;EXTERNAL LINK SIGN;So;0;ON;;;;;N;;;;;

> 2B*x*2;INTERNAL LINK SIGN;So;0;ON;;;;;N;;;;;

IV. Contemporary marking of external and internal links

As there exists no standardized characters now, site authors use to create little graphic which they use for the external and internal link symbols throughout their site. The existence of standardized characters as proposed would free the site authors from having to reinvent the wheel for that purpose.

As site authors are free to design those symbols (in absence of any standard), some diverging designs are found (see Appendix), but there is a clear evidence of an evolving convergence towards a rectangle with an arrow pointing outward from it, resp. a box with an arrow inside.

Sites which use an arbitrary design for an external link symbol use to explain it, while some sites using a rectangle with an arrow pointing outward design do not in some cases, thus showing more confidence in the understandability of the symbol.

This is evidence that a sufficiently universally recognized symbol has emerged.

V. Acknowledgements

Drafts of an earlier version of this proposal (containing the EXTERNAL LINK SIGN only) were discussed on the public Unicode E-mail list in 2006. Thanks to all participants. Many hints from there are incorporated into this proposal.

VI. SPECIMENS

1. External link sign, collected 2006-07-21 from:
    http://en.wikipedia.org/wiki/Unicode

    ### External links

    - The Unicode Consortium ⊡
        - Unicode versions: 3.1 ⊡, 3.2 ⊡, 4.0 ⊡, 4.0.1 ⊡, 4.1 ⊡ 5.0.0 ⊡ (beta) ⊡
    - Unicode Character Search ⊡
        - new characters ⊡, scripts ⊡ and characters and scripts under investigation ⊡
        - Code Charts ⊡ (PDF files)

2. External and internal link sign, collected 2006-07-21 from:
    http://www.heise.de/tp/r4/artikel/23/23144/1.html (www.heise.de is the site of a German publisher, publishing e.g. the German computer magazine "c't").

    Die Region gilt als Wiege der Menschheit, denn in der Afar-Senke fanden sich Fossilien verschiedener Menschen- und Hominidenarten (vgl. ⊠ Neue Affenmenschen), darunter auch die berühmte, bereits aufrecht schreitende und 3,5 Millionen Jahre alte ⊡ Lucy.

    Diese Tiefebene in Äthiopien ist Teil des afrikanischen Grabens, dort stoßen die arabische, die nubische und die somalische Platte aufeinander. Eine unruhige und vulkanisch aktive Gegend, in der sich seit 30 Millionen Jahren immer wieder tiefe Gräben auftun. Das Rote Meer entstand so, als sich die arabische Erdplatte vom afrikanischen Kontinent entfernte (vgl. ⊡ Plattentektonik).

3. External link sign, collected 2006-07-21 from:
    http://www.signographie.de/cms/front_content.php?idcat=1&idart=57

    Herunterladen und Informationen: ⊡ http://www.europatastatur.de

4. External link sign, collected 2006-07-21 from:
http://hackage.haskell.org/trac/ghc

- Pick an easy bug report or task to work on
- Grab the latest sources
- ⇨ Set up your build tree
- Useful information about GHC's architecture is in the ⇨ commentary
- Hack away... ask questions on ⇨ the cvs-ghc mailing list
- Submit patches back, using darcs send is the recommended way.

5. External link sign, collected 2006-07-21 from:
http://investor.invescoperpetual.co.uk/invesco/default/page/literature/request/list_single_copy/0,2879,
32524157_34698815,00.html

Adobe®
Visit Adobe.com
↗ Access Adobe online conversion tools

Get Adobe Reader
↗ Install the latest Adobe Reader here

When using an external link - ↗ - you will be leaving our website. Any
views and opinions expressed subsequently are not ours.

6. External link sign (in the center of the second last line), collected 2010-07-15 from:
http://www.alternate.de/html/shop/productDetails.html?artno=IMEL3M&checksum=HTRaGCGIy8pmAE
Jzbk3CYVsBDiiBv6EIlnFNjO2RFDk%3D (a product description site of a German hardware supplier).

Ausführliche Details	Zubehör	Bewertungen	Preisentwicklung

Für Anwender die Leistung, Kapazität und hohe Qualität verlangen, ist der USB-Stick JetFlash 620 von
Transcend die erste Wahl. Dank seiner fortschrittlichen Dual-Channel-Architektur bietet der JetFlash
schnelle Datentransferraten von bis zu 32 MB/s (lesen). Die mitgelieferten JetFlash-elite-
Data-Management-Tools bieten Komfortfunktionen wie z.B. E-Mail-Zugriff und Browser-Favoriten sowie

VII. Appendix: Google image search for "external link icon"

A Google image search on 2006-07-30 yielded:

26 "external link signs" from different sources are in fact listed here, not counting the last symbol in the first row (as it serves another purpose), the duplicate in the second last row, and the trademark in the center of the last row.

Of these, 16 resemble (more or less) the "box with arrow pointing out of it" pattern.
7 other ones resemble a "circle or world with arrow pointing from it", "boxed arrow", or a simple "pointing symbol" pattern. Only three follow a completely different pattern.

<table>
<tr><td colspan="2" align="center">ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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.</td></tr>
</table>

A. Administrative

1. **Title:**	*Proposal to encode signs for external and internal links in the UCS*
2. Requester's name:	*Karl Pentzlin*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*2012-05-03*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*Yes*
(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: *Yes*
        Name of the existing block: *Miscellaneous Symbols and Arrows*

2. Number of characters in proposal: *2*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	X	B.1-Specialized (small collection)	B.2-Specialized (large collection)
C-Major extinct		D-Attested extinct	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols	

4. Is a repertoire including character names provided?   *Yes*
    a. If YES, are the names in accordance with the "character naming guidelines"
        in Annex L of P&P document?   *Yes*
    b. Are the character shapes attached in a legible form suitable for review?   *Yes*

5. Fonts related:
    a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
        The author – font downloadable at http://www.pentzlin.com/LinkSigns.ttf
    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
        The font is released into the Public Domain by the author

6. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *Yes*
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
    of proposed characters attached?

7. Special encoding issues: *Yes*
    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)?   *n/a*

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? *Partially*
   If YES explain *No decision was made on the original proposal (L2/06-268)*
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.)? *Yes*
   If YES, with whom? *The author himself is a member of the user community*
   If YES, available relevant documents: *see text*
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference: *see text*
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference: *see text*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *Ubiquitous on Internet pages*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? *Yes*
   If YES, is a rationale provided? *Yes*
   If YES, reference: *To keep them in line with related arrow-like symbols*
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? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
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? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *No*
    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:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: