L2/18-303

Revised EXTERNAL LINK Proposal

To: Unicode Consortium
Author: Devin Bayer <dev@doubly.so>
Date: 2018-09-18

Introduction

EXTERNAL LINK (🔗) is primarily used in hypertext to mark links pointing outside of the current domain or document. As opposed to internal links, which typically point to another section of the current domain or document. It's not used for every link but only when needed in context.

🔗 is a popular character, available in many symbol-oriented icon packs and fonts. It has been proposed twice before[1] but neither was able to overcome the *plain text* criteria applied to new characters. I won't repeat the parts from earlier proposals which seamed adequate, but instead directly address the justifications for previous rejections.

Please see the first proposal (L2/06-268 🔗) for examples collected from the wild, demonstrating the symbol is widely used.

Plain Text Objection

Accord to the UTC Minutes[2], the following rational was given for rejecting the symbol:

> The UTC rejected the proposals to add "external link sign", most recently in L2/12-169. It is unclear that the entity in question is actually an element of plain text, given the inevitable connection to its function in linking to other documents, and thus its coexistence with markup for links. Furthermore, the existing widespread practice of representing this sign on web pages using images (often specified via CSS styles) would be unlikely to benefit from attempting to encode a character for this image. (This notice of non-approval should not be construed as precluding alternate proposals which might propose encoding a simple shape-based symbol or symbols similar in appearance to the images used for external link signs, should an appropriate plain-text argument for the need to encode such a simple graphic symbol be forthcoming.)

Indeed the external link character is primarily useful in hypertext where markup is available, but as the following examples will show, it doesn't follow that a graphical image is a suitable substitution for a plain-text character. Hopefully, I can demonstrate that the symbol carries useful semantics required in non-graphical contexts.

---

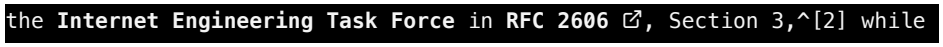
[1] See L2/06-268 🔗, 2006 and L2/12-169 🔗, 2012

[2] : Document 131-C26, May 10, 2012

Example 1: Text-Based Browsers

Browsers that operate in consoles where only text is available are still in occasional use and actively developed as of 2018.[3] Websites which make ample use of ↗ in graphical contexts are not able to indicate external links in the console. Take the following screenshot of Wikipedia from the *links* browser:

The link to RFC 2606 is an external link to tools.ietf.org but is indistinguishable from a link to another article. With EXTERNAL LINK (↗) as a character, it would be possible to increase the clarity with something like the following:



[3] https://en.wikipedia.org/wiki/Links_%28web_browser%29 ↗

Example 2: Text-Based Websites

Some sites remain text-only or provide text-only alternatives. The advantages of text-only, especially reducing server load, are most important for major news outlets to handle traffic spikes after significant events. That's why NPR maintains https://text.npr.org/ and CNN maintains http://lite.cnn.io/en.

This example shows how an article on NPR's text-only site could benefit from EXTERNAL LINK ().

> Text-Only NPR.org (go to full version)
>
> Home
>
> Something happened dipiscing magna sed dolor elit
>
> By Somebody
>
> Adipiscing magna sed dolor elit. Praesent eleifend dignissim arcu, at eleifend sapien imperdiet ac. Aliquam erat volutpat. Praesent urna nisi, fringila lorem et vehicula lacinia quam. Integer sollicitudin mauris nec lorem luctus ultrices statement by Adipiscing magna sed dolor elit.
>
> - Contact Us
> - Terms of Use
> - Permissions
> - Privacy Policy
>
> © NPR

Example 3: Web Accessibility

There have been requests[4] for a standard external link symbol for the purpose of increasing web accessibility. The current practice of using non-standard images carry no semantics, so can't be interpreted by screen readers or other automated tools.

[4] https://stackoverflow.com/questions/1899772/new-window-icon-for-web-accessibility

Example 4: Simpler Graphical Usage

Although the rejection was accurate that in purely graphical contexts the character is not *necessary* because images can be substituted, images are significantly more complex.

The best-practice method to include custom icons in modern browsers is with inline SVG. This has the advantage that no external resources need be packaged and it scales like vector fonts. Without the EXTERNAL LINK (), this is the amount of code needed to use this symbol:

```
<a href="...">
      <svg style="display: inline-block; width: 1em; height: 1em;" viewbox="0
      0 48 48">
      <path d="M36 24c-1.2 0-2 0.8-2 2v12c0 1.2-0.8 2-2 2h-22c-1.2
      0-2-0.8-2-2v-22c0-1.2 0.8-2 2-2h12c1.2 0 2-0.8 2-2s-0.8-2-2-2h-12c-3.4
      0-6 2.6-6 6v22c0 3.4 2.6 6 6 6h22c3.4 0 6-2.6
      6-6v-12c0-1.2-0.8-2-2-2z"></path>
      <path d="M43.8 5.2c-0.2-0.4-0.6-0.8-1-1-0.2-0.2-0.6-0.2-0.8-0.2h-12c-1.2
      0-2 0.8-2 2s0.8 2 2 2h7.2l-18.6 18.6c-0.8 0.8-0.8 2 0 2.8 0.4 0.4 0.8
      0.6 1.4 0.6s1-0.2 1.4-0.6l18.6-18.6v7.2c0 1.2 0.8 2 2 2s2-0.8
      2-2v-12c0-0.2 0-0.6-0.2-0.8z"></path>
      </svg>
</a>
```

With the character available this changes to:

```
<a href="...">&#x1F517;</a>
```

Hopefully, this is a convincing demonstration of how the rejection missed the mark on this point:

> the existing widespread practice of representing this sign on web pages using images (often specified via CSS styles) would be unlikely to benefit from attempting to encode a character for this image.

Character Properties

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

```
2Bx1;EXTERNAL LINK;So;0;ON;;;;;N;;;;;
```

Symbol Encoding Criteria

The following criteria[5] are considered for encoding symbols.

Pros (6 apply)

+ is typically used as part of computer applications (e.g. CAD symbols)
+ has well defined user community / usage
+ always occurs together with text or numbers (unit, currency, estimated)

  must be searchable or indexable

  is customarily used in tabular lists as shorthand for characteristics (e.g. check mark, maru etc.)

  is part of a notational system
+ has well-defined semantics
+ has semantics that lend themselves to computer processing

  completes a class of symbols already in the standard
+ is letterlike (i.e. should vary with the surrounding font style)

Cons (0 apply)

the symbol is primarily used freestanding (traffic signs)

the notational system is not widely used on computers (dance notation, traffic signs)

the symbol is part of a set undergoing rapid changes

the symbol is trademarked (unless requested by the owner) (logos, Der grüne Punkt, CE symbol, UL symbol, etc)

is purely decorative

it's ok to ignore its identity in processing

font shifting is the preferred access and the user community is happy with it (logos, etc.)

[5] https://www.unicode.org/pending/symbol-guidelines.html

Drawing the Symbol

In its essence, EXTERNAL LINK (🔗) is a box with an arrow starting from its center, pointing pointing “north-east”. The arrowhead is outside. The box has a gap in the corner where the arrow crosses it. The proposed character is not part of any script and the precise form of their drawing is not critical.

It’s probably more appropriate to reverse the direction of the symbol for right-to-left scripts, since the arrow should be pointing away from the link name. Take for reference the Arabic Wikipedia:

Conclusion

EXTERNAL LINK (🔗) is a popular symbol used to visually mark hyperlinks which point to external sources. It’s unusual for Unicode in that it’s only relevant in contexts where links are possible (hypertext). However, hypertext does not imply that anything except plain-text characters are available. At least the first 2 examples in this proposal show contexts where there is no substitute. The 3rd example shows that even in graphical contexts, having the character standardised is more convenient.

Due to all the benefits of a standardised 🔗 character outlined in the this proposal and those from 2006 / 2012, I recommend Unicode include it in the next version.

References

L2/06-268 🔗 - Original Proposal for the EXTERNAL LINK (2006)

L2/12-169 🔗 - Proposal to encode signs for external and internal links in the UCS (2012)

L2/12-143 🔗 - Comment on External Link Sign proposal handling

L2/12-177 🔗 - Feedback on the proposal L2/12-169 for Link Signs

🔗 - LINK SYMBOL (U+1F517)

🗗 - OVERLAP (U+1F5D7)

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[6]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Revised EXTERNAL LINK Proposal*
2. Requester's name: *Devin Bayer*
3. Requester type: *Individual*
4. Submission date: *2018-09-18*
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:
      Name of the existing block: *Miscellaneous Symbols and Arrows*
2. Number of characters in proposal: *1*
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?
   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?
   Feather is one example icon set, available under an open-source license
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   https://feathericons.com
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? *Yes*
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)? *No*

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.

[6] Form number: N4502-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? *Yes*
   If YES explain *Proposals in 2006 and 2012 were rejected due to lack of plain-text usage.*
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.)? *No*
   If YES, with whom?
   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*
   Reference: *This document*
4. The context of use for the proposed characters (type of use; common or rare) *average*
   Reference: *This document*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *Most notably, Wikipedia. Also see L2/06-268*
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: *Keep with other Miscellaneous 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: