Internalization of 'ftp' URIs
draft-yevstifeyev-ftp-iri-00

Abstract

This document discusses internalization of 'ftp' Uniform Resource Identifiers (URIs), which are used to reference resources accessible via File Transfer Protocol (FTP). It updates RFC-tbd.

[NOTE TO RFC EDITOR: RFC-tbd refers to the [FTP-URI]. Please replace "tbd" when [FTP-URI] becomes an RFC with its number.]

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect
1. Introduction

RFC-tbd [FTP-URI] defined the Uniform Resource Identifier (URI) scheme used to refer to the resources accessible via File Transfer Protocol (FTP) [RFC0959]. RFC 3986 [RFC3986] allows a limited range of characters to be present in URIs (particularly, only characters in ASCII range [ASCII]). RFC 3987 [RFC3987] introduced a new protocol element - Internationalized Resource Identifiers (IRIs) - which supplements URIs by extending the allowed range of characters to all Universal Character Set (UCS) characters. UCS is defined by [ISO.10646] and [UNICODE] simultaneously, and includes the characters which are able to represent almost any modern script and language. See RFC 3536bis [RFC3536bis] for more information.

This document discusses the internalization of ‘ftp’ URIs and, correspondingly, defines ‘ftp’ IRIs. It updates RFC-tbd [FTP-URI].

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "MAY", "OPTIONAL", "MAY NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MUST BE", and "SHALL BE" in this document have the meanings defined in BCP 14, and are used throughout this document.

Table of Contents

1. Introduction ................................................. 2
   1.1. Terminology .......................................... 2
2. Internalized ‘ftp’ Resource Identifiers (RIs) .................. 3
   2.1. ‘ftp’ IRIs ............................................ 3
   2.2. UCS Characters in ‘ftp’ URIs ............................ 4
3. Handling Internalized ‘ftp’ RIs ................................ 4
   3.1. Internalized <host> Part in ‘ftp’ URIs and <ihost> Part in ‘ftp’ IRIs .................................. 4
   3.2. The <user-pass> Part in ‘ftp’ URIs and IRIs............... 4
   3.3. The <ftp-path> Part in ‘ftp’ URIs and <iftp-path> in ‘ftp’ IRIs ........................................ 5
4. Internalization of Actual Data Interchange ........................ 5
5. Mapping ‘ftp’ IRIs to ‘ftp’ URIs ................................ 5
6. Security Considerations ........................................ 5
7. IANA Considerations .......................................... 5
8. References .................................................... 6
   8.1. Normative References .................................... 6
   8.2. Informative References ................................... 7
Appendix A. Acknowledgments ........................................ 7
Authors’ Addresses ............................................... 7
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

The reader is assumed to be familiar with the terminology of RFC 3987 [RFC3987], RFC 5890 [RFC5890], RFC 3536bis [RFC3536bis] and RFC-tbd [FTP-URI].

In this document Resource Identifier (RI) refers to either URI or IRI.

2. Internalized ‘ftp’ Resource Identifiers (RIs)

The internalized ‘ftp’ Resource Identifier is meant to be a ‘ftp’ IRI and ‘ftp’ URI with UCS characters. The subsequent sections discusses how to use UCS with ‘ftp’ RIs.

2.1. ‘ftp’ IRIs

As mentioned before, IRIs provide a way to include UCS characters in the resource identifier preserving the functions of URI. Unlike URIs, which also may contain UCS characters, which occur percent-encoded, though (see Section 2.2), IRIs represent UCS characters directly, by first encoding them using UTF-8 [RFC3629]. This makes use of IRIs by humans much easier. This section defines the valid syntax of ‘ftp’ IRIs.

The ‘ftp’ IRI takes the form of <ftp-iri> rule below, defined using ABNF [RFC5234]:

```
ftp-iri = "ftp:" ftp-ihier-part
ftp-ihier-part = "//" [ user-pass "@" ] ihost-port [ iftp-path ]
user-pass = <as defined in RFC-tbd [FTP-URI]>
  ; not internalized - see below
ihost-port = ihost [ ":" port ]
iftp-path = [ icwd-part ] "/" ilast-segment [ typecode-part ]
icwd-part = *( "/" icwd )
icwd = isegment-nsc
last-segment = isegment-nsc
isegment-nsc = *ipchar-nsc
ipchar-nsc = iunreserved / pct-encoded / sub-delims-nsc / ":"
  / ":@",
sub-delims-nsc = "!" / "$" / ":" / ":" / "@" / "#" / "" / "" / "" / ":" / ":" / ":" / ":@",
typecode-part = <as defined in RFC-tbd [FTP-URI]>
```
where the rules <ihost> and <iunreserved> are defined in RFC 3987 [RFC3987], <pct-encoded> and <port> - in RFC 3986 [RFC3986], and <user-pass> and <typecode-part> - in RFC-tbd [FTP-URI].

2.2. UCS Characters in ‘ftp’ URIs

Valid ‘ftp’ URIs MAY contain UCS characters in their <host> and <ftp-path> parts, as mentioned in Section 3.3 of RFC-tbd and Section 2.5 of RFC 3986. In order to use such character in one of these parts, it SHALL first be encoded with UTF-8 [RFC3629]. The resulting sequence of octets SHALL be examined to conclude whether some octets match corresponding ASCII characters. If one does, and such character is allowed in a particular part of ‘ftp’ URI, it SHALL be presented in the URI directly; otherwise, the octet SHALL be represented percent-encoded.

3. Handling Internalized ‘ftp’ URIs

This section defines the rules which concern handling internalized ‘ftp’ URIs. The valid syntax for ‘ftp’ URIs is defined above; how to include UCS characters is also discussed above.

3.1. Internalized <host> Part in ‘ftp’ URIs and <ihost> Part in ‘ftp’ IRIs

The <host> part in ‘ftp’ URIs and <ihost> part in ‘ftp’ IRIs may contain internalized strings, with UCS characters being percent-encoded and displayed directly, respectively.

As Domain Name System (DNS) does not allow the UTF-8 encoded data in its interchange, limiting the allowed characters range to ASCII [ASCII], the usual procedure of UTF-8 transformation is insufficient here. Hence, in order to make up the valid domain name for lookup and further processing the following step SHALL be applied:

(1) represent the domain name which contains UCS characters as an octets stream in UTF-8 [RFC3629];

(2) apply the algorithm for IDN lookup defined in Section 5 of RFC 5891 [RFC5891].

The received A-label SHALL also be used with FTP HOST command [I-D ietf-ftpext2-hosts], sent when establishing FTP connection per Section 3.2.1 of RFC-tbd [FTP-URI].

3.2. The <user-pass> Part in ‘ftp’ URIs and IRIs

The <user-pass> part of ‘ftp’ IRIs is not internalized. FTP does not
allow the UCS characters to be present in user names or passwords, and does not allow to enable such feature. Correspondingly, when processing ’ftp’ RIs, either URIs or IRIs, the <user-pass> may contain characters only allowed by RFC-tbd [FTP-URI].

3.3. The <ftp-path> Part in ’ftp’ URIs and <iftp-path> in ’ftp’ IRIs

The <ftp-path> and <iftp-path> parts allow to include UCS characters in FTP pathnames present in URIs and IRIs, respectively.

In order to successfully process an internalized FTP pathname, a client prior to its processing SHALL examine the server’s response to the FEAT command [RFC2389], issued upon authentication per Section 3.2.2 of RFC-tbd [FTP-URI]. If one of the lines of the response is "UTF8", the server supports UTF-8 encoded pathnames [RFC2640]. Otherwise, if there is no such line, or the server does not support the FEAT mechanism, the contrary SHALL be assumed.

Should it be determined that server supports UTF-8 encoded pathnames, the internalized pathname parts SHALL be encoded with UTF-8 [RFC3629] and then transmitted as an arguments to the corresponding FTP commands as UTF-8 octets stream.

4. Internalization of Actual Data Interchange

’ftp’ RIs may refer to the file which contains the internalized data. When transmitting such file over data connection, it should be in Net-Unicode format [RFC5198]. In order to indicate this, the <typecode> equal to "u" [I-D ietf-ftpext2-typeu] SHALL be set in the ’ftp’ RI.

5. Mapping ’ftp’ IRIs to ’ftp’ URIs

’ftp’ IRIs are subject to the rules of Section 3.1 of RFC 3987 with respect to their mapping to ’ftp’ URIs.

6. Security Considerations

Security considerations for usage of IRIs are discussed in Section 8 of RFC 3987 [RFC3987]; for usage of Internalized Domain Names – in RFC 5890 [RFC5890]. This document adds nothing to those security considerations.

7. IANA Considerations

IANA is asked to list this document as an additional reference to ’ftp’ URI scheme registration.
8. References

8.1. Normative References


8.2. Informative References

[I-D ietf-ftpext2-hosts]


[RFC3536bis]

Appendix A. Acknowledgments

I’d like to thank <TBD> for their input to this document.

Authors’ Addresses

Mykyta Yevstifeyev
8 Kuzovkov St., Apt 25
Kotovsk
Ukraine

EMail: evnikita2@gmail.com